Finite Geometry & Friends

18-22 September 2023
Vrije Universiteit Brussel
Brussels
Belgium

Course notes

DOCTORAL SCHOOL OF NATURAL SCIENCES & (BIOSCIENCE) ENGINEERING

The Institute of Combinatorics and its Applications

Summer School

Finite Geometry and Friends

Vrije Universiteit Brussel

Brussels, Belgium

September 18–22, 2023

Lecture notes

Krystal Guo
University of Amsterdam

Anna-Lena Horlemann
University of St. Gallen

Valentina Pepe
Sapienza University of Rome

John Sheekey
University College Dublin

Contents

Preface **v**

I Graph Isomorphism and Matrix Algebras
(Krystal Guo) **1**

II An Introduction to Code-Based Cryptography
(Anna-Lena Horlemann) **35**

III Points of Algebraic Varieties in generic position
(Valentina Pepe) **69**

IV The Geometry of Semifields
(John Sheekey) **97**

Preface

Dear participants,

we are very pleased to welcome you at the second edition of the summer school[1]

Finite Geometry and Friends,

held at the Vrije Universiteit Brussel (Belgium) from September 18$^{\text{th}}$ to 22$^{\text{nd}}$, 2023. Four young, yet established researchers will each deliver four hours of lectures on their topics, supplemented by two hours of exercise sessions. Each of them has written a set of notes to accompany the lectures.

First up, we have Krystal Guo, from the University of Amsterdam. She will discuss algorithms to determine whether or not two graphs are isomorphic. These algorithms are closely related to algebras generated by matrices attached to the graphs. Therefore, Krystal will introduce us to distance regular graphs, which are graphs generating very well-behaved matrix algebras. This is should be of interest to any finite geometer, since a lot of the classical families of distance regular graphs are constructed from geometries.

Next in line, we have Anna-Lena Horlemann, from the University of St. Gallen. She will introduce us to code-based cryptography. This is one of the leading avenues being explored in post-quantum cryptography. While finite geometry is known for its connection to coding theory rather than cryptography, code-based cryptography builds a bridge between the areas of coding theory and cryptography, and is hence an invitation for the finite geometer to cross over into the realm of cryptography.

The third lecturer which we have the pleasure of introducing is Valentina Pepe from Sapienza University of Rome. The topic of her lectures are large sets of points in a projective space, not containing a small number of linearly dependent points. Such point sets are usually constructed from algebraic varieties, and in

[1] This year's edition is technically both a summer and an autumn school.

some cases algebraic varieties are the only possible way to construct these point sets. This topic has a well-known connection to coding theory, but is also relevant in extremal graph theory.

Last but not least, John Sheekey, from University College Dublin, will share with us his knowledge on semifields, and their links to various geometric objects. Semifields are certain algebraic structures, satisfying less restrictive requirements than fields. They lead to the construction of spreads in projective spaces, which in turn give rise to constructions of axiomatic affine and projective planes. In addition, semifields are used in of coding theory, more specifically rank-metric codes, and also play an important role in the study of tensors. The versatility of these objects provides ample motivation for their exploration.

These lecture notes show a varied range of topics with interesting links to finite geometry. We have the honor of having four top-level researchers sharing their expertise with us. We hope that you, the participants of the summer school, enjoy being introduced to these topics, find inspiration for future research, and make some friends along the way.

Sincerely,

the organizers,

Sam Adriaensen

Jan De Beule

Leen Demuys

Jonathan Mannaert

Sam Mattheus

Part I

Graph Isomorphism and Matrix Algebras

Krystal Guo

Korteweg-de Vries Instituut for Mathematics,
University of Amsterdam,
P.O. Box 94248,
1090 GE Amsterdam,
The Netherlands

email: guo.krystal@gmail.com

Contents

Preface **5**

1 Introduction **7**
1.1 Cospectral graphs 8
1.2 Graph Isomorphism algorithms 9
1.3 Distinguishing SRGs with spectra 9

2 Distance-regular graphs **13**
2.1 Examples 14

3 Matrix algebras **19**
3.1 Association schemes 20
3.2 Cospectrality 21
3.3 Weak and strong isomorphism 22

4 Weisfeiler-Lehmann algorithm **23**
4.1 Cai-Fürer-Immerman graphs 24
4.2 Connections to Weisfeiler-Lehman 25

5 Extensions of matrix algebras **29**
5.1 Equivalence of extensions 29
5.2 Triply regular graphs 30
5.3 Other matrix algebras 31

Bibliography **33**

Preface

In these notes, we embark on an exploration approaching graph isomorphism using algebraic graph theory. Our primary goal is to establish a framework to prove the cospectrality of matrices associated with graphs, by showing the matrices belong to "cospectral" matrix algebras – part of the notes are to make the concept of cospectral matrix algebras rigourous.

We begin with a discussion of the Graph Isomorphism problem and various graph invariants that have been proposed to be complete graph invariants for the class of strongly regular graphs, to which we give a brief introduction in Chapter 2. We proceed to define cellular algebras and weak and strong isomorphisms between them. After a short excursion to Weifeiler-Lehman algorithm, we build on the matrix algebras to introduce extensions in Chapter 5 and how they relate to the graph invariants in Chapter 1.

Chapter 1

Introduction

Given two graphs, the decision problem of deciding whether or not they are isomorphic is known as the *Graph Isomorphism* problem. It is one of the few problems which are not known to be NP-complete and also not known to have a polynomial-time algorithm, and lives in its own complexity class (GI). One possible direction of research is in algorithms, some of which we will address in Section 1.2. The direction we will focus on is that of graph invariants.

A graph property is a graph invariant if it is invariant under all isomorphism of the graph. For example, while the adjacency matrix of the graph is not a graph invariant, the eigenvalues of the adjacency matrix are. Other examples of graph invariants include the number of vertices, the number of edges, the clique number and the co-clique number. A graph invariant is said to be complete, if the equality of the invariants implies the isomorphism of the graphs.

For example, we can write an adjacency matrix as a string of `0`s and `1`s. For an isomorphism class, we can take all such strings coming from the adjacency matrices of graphs therein and extract the lexicographically smallest string. This string is a complete graph invariant; two graphs are isomorphic if and only if this process results in the same string. In SageMath, this is (essentially) implemented by the `graph6_string` of the `canonical_label()` of a graph.

By the above discussion over the complexity, we know that there is no polynomial-time computable complete graph invariant for the class of all graphs. This has led to a pursuit of complete graph invariant, especially those of a spectral flavour, for restricted classes of graphs, like strongly regular graphs. We will explore the literature more fully in Section 1.3.

Eigenvalues of graphs are graph invariants with respect to many choices of matrices. One can ask for which classes of graphs are they complete invariants and how we might prove or disprove such a statement? Answer: Matrix algebras.

Main topics:

- strongly regular graphs
- Bose-Mesner algebras of graphs
- cellular algebras
- weak and strong isomorphisms of matrix algebras
- Weisfeiler-Lehman algorithm and its associated hierarchy of matrix algebras
- Terwilliger algebras, Jaeger algebras

In the remainder of this chapter, we will give some context and motivation for our study, with a brief overview of Graph Isomorphism algorithms and various matrices whose spectra may (or may not) distinguish strongly regular graphs.

1.1 Cospectral graphs

No discussion of distinguishing graphs using spectra would be complete without a brief note on cospectrality with respect to the adjacency matrix. The *adjacency matrix* of a graph X, denoted $A(X)$ or A when the context is clear, is the matrix with rows and columns indexed by the vertices of X, where

$$A(u,v) = \begin{cases} 1, & \text{if } uv \in E(X); \\ 0, & \text{otherwise.} \end{cases}$$

Since we will are mostly interested in regular graphs, the adjacency spectrum and various Laplacian spectra will be equivalent, for the purposes of determining graphs.

Note that, unless otherwise specified, we refer to the spectrum of the adjacency matrix of a graph as its *spectrum* and two graphs are *cospectral* if their adjacency matrices are. For the other spectral invariants in these notes, we either define a (usually larger) auxiliary graph and take its adjacency matrix, or we define a different matrix from adjacencies between various substructure of the graph, such as edges or directed edges.

A major open problem in spectral graph theory is the following conjecture, which has been attributed to Haemers and also to Babai:

> Conjecture: As n grows large, the proportion of graphs on n vertices with no cospectral mate approaches 1.

For cospectrality of graphs on up to 12 vertices, see [7]. In the other direction, any two strongly regular graphs with the same parameter set are cospectral; the smallest such pair is the Schrikhande graph and the rook graph on 16 vertices.

1.2 Graph Isomorphism algorithms

In 1980, Babai gave an algorithm to assign a canonical labeling to the vertices of a strongly regular graph whose running time is $o(\exp(2n^{1/2}\log^2 n))$. This was improved by Spielman in [22], by giving an algorithm for testing isomorphism of strongly regular graphs in $n^{O(n^{1/3}\log n)}$. At the time, the best known algorithms for testing graph isomorphism of general graph was $2^{\sqrt{O(n\log n)}}$. This gave an intuition that strongly regular graphs might be a class of graphs where graph isomorphism is an easier problem than in general graphs. In some sense, the proposed algorithms in Section 1.3 are attempts of compute complete graph invariants for strongly regular graphs, and are motivated by this intuition.

In 2015, László Babai announced a quasi-polynomial time algorithm for Graph Isomorphism [4] and the extended abstract [5] won the best paper award at STOC '16. In 2017, Harald Helfgott pointed out an error in the analysis that indicated that the algorithm runs in sub-exponential time, as opposed to quasi-polynomial. Soon after, Babai modified the proof with the claim that it achieves the original claimed result and posted the fix for the flaw. As it stands (to the best of my knowledge as of August 23, 2023), the arXiv paper, which has over 800 citations, has not been updated to include this and other fixes and Babai's website[3] reports that this result has not been completely peer-reviewed. The progression of this paper and Babai's framework has inspired many lines of research and rejuvenated the area.

1.3 Distinguishing SRGs with spectra

One can associate with a graph many matrices; there have been many proposed algorithms for graph isomorphism in the class of strongly regular graphs which use various different matrices. We will look at several of these matrices in these notes. The examples include symmetric squares, symmetric powers (matrix indexed by k-subsets of vertices, with a 1 when two k-subsets have an edge as their symmetric difference), and also various quantum-walk inspired matrices.

1.3.1 Symmetric powers

The *symmetric kth power* $X^{\{k\}}$ of a graph X is given as follows: the vertices are the k-subsets of $V(X)$, and two k-subsets are adjacent if and only if their symmetric difference is an edge of X. These are also known as *k-token graphs* in the literature. Graph invariants of the symmetric powers of X are also graph invariants of X itself; in particular, the spectra of the symmetric powers of X are graph invariants and are computable in polynomial time provided that k is a constant. One can find many pairs of cospectral graphs whose symmetric squares

are not symmetric. On the other hand, Audenaert et al. [2] show that if X and Y are cospectral strongly-regular graphs then $X^{\{2\}}$ and $Y^{\{2\}}$ are cospectral, but they leave the possibility that there is some k where the spectra of $X^{\{k\}}$ and $Y^{\{k\}}$ can distinguish any two strongly regular graphs X, Y.

Alzaga, Iglesias, Pignol [1] show that if the $2k$-dimensional Weisfeiler–Lehman method fails to distinguish two given graphs, then their kth symmetric powers are cospectral. Since it is well-known, that there are pairs of non-isomorphic n-vertex graphs which are not distinguished by the k-dimensional Weisfeiler-Lehman method, [1] shows that, for each k, there are pairs of non-isomorphic n-vertex graphs with cospectral k-th (symmetric) powers. Independently, Barghi and Ponomarenko [6] show that given a positive integer m, there exist infinitely many pairs of non-isomorphic graphs with cospectral m-th symmetric powers, using extensions of coherent configurations, which we will look at in Chapter 5.

We note that neither of the examples in [1, 6] are strongly cospectral. In particular, we note the following open problem:

> Open Problem: Does there exists a fixed m such that any two strongly regular graphs are isomorphic if and only if their m-th symmetric powers are cospectral?

The result of [2] says that m must be strictly greater than 2. Some computations have been done (in [2]), but I am not aware of any pairs of strongly regular graphs with cospectral symmetric cubes.

There is a related invariant proposed by Gamble, Friesen, Zhou, Joynt and Coppersmith [12]; this invariant takes the entries of the k-particle transition matrix as a list. A special case of this transition matrix corresponds to the symmetric square of the graph. Though it does not use the spectrum of this matrix, we will discuss it as it is proposed as a complete invariant for the class of strongly regular graphs for which no counterexample is known. The matrix in question is the transition matrix of a quantum process involving k particles, exhibiting characteristic of Fermions or Bosons.

Smith showed that, for any k, there exists a pair of non-isomorphic graphs that are not distinguished by the k-Boson invariant. Strongly regular graphs which are not distinguished by this procedure are yet unknown.

1.3.2 Discrete-time quantum walk

The discrete-time quantum walk is a quantum process on a graph. The important aspect (for us) is that it has a transition matrix with combinatorial information.

We take a graph and replace every edge with two oppositely oriented edges. We can define the heads incidence matrix; it is a matrix D_h indexed by vertices and directed edges ($v \times 2e$) where $D_h(u, e) = 1$ if u is the head of e. Similarly, we have

the tails incidence matrix D_t. Let P be the $2e \times 2e$ permutation matrix which takes each arc to the oppositely oriented arc. Thus we have,

$$D_h P = D_t, D_t D_h^T = A = D_h D_t^T.$$

The transition matrix of the DT quantum walk of a k-regular graph is

$$U = \frac{2}{k} D_t^T D_h - P.$$

Emms et al. 2006 [10] proposed to use the following to distinguish SRGS: obtain S is obtained from U^3 by replacing all positive entries with 1 and all other entries with 0, then output the spectra of S. The operation of replacing all positive entries of a matrix M with 1 and all other entries with 0 is called taking the *positive support* of M and denoted $S^+(M)$.

This does distinguish many pairs of strongly regular graphs. For example, there exists non-isomorphic Paley and Peisert graphs with parameters $(49, 20, 11, 12)$, which are distinguished by this process. For odd q, $GQ(q, q)$ is not isomorphic to its dual and they are distinguished by their for small q. In [14], we did finally find that this does not distinguish a specific pair of GQs; there exists two $GQ(5, 25)$ which are not distinguished. The matrix is 98280×98280, dense, and not symmetric.

Chapter 2

Distance-regular graphs

Connected strongly regular graphs are distance-regular graphs of diameter 2. Since the machinery with the matrix algebras that we will look at for strongly regular graphs will carry over for distance-regular with little adaption, we will give a brief introduction to distance-regular graphs, from the point of view of group actions on graphs, in this chapter. We defer to [15, 13] for further background.

We can consider automorphisms of graphs; ϕ is an *automorphism* of a graph X if ϕ is a permutation of the vertices of X such that edges are mapped to edges. The set of all automorphism of a graph forms a group, $\mathrm{Aut}(X)$. Equivalently, $\mathrm{Aut}(X)$ is the group of permutation matrices in $V(X) \times V(X)$ which commute with the adjacency matrix $A(X)$. We can consider group action on our graph X; these will be homomorphism from group G to $\mathrm{Aut}(X)$.

Since we have defined automorphisms as permutation of the vertices, it is natural to think of the group acting upon the vertices of the graph. However, we may also consider the group action upon subsets of vertices. And $g \in Aut(X)$ also acts on the pairs of vertices as follows:

$$g(x, y) = (g(x), g(y)).$$

The *orbitals* of the graph are the orbits of $V(X) \times V(X)$ under the action of the automorphism group. Since we cannot map (x, y) at distance i from each other to (u, v) at distance $j \neq i$, there are at least $d + 1$ orbitals, where d is the diameter of the graph. If there are exactly $d + 1$ orbitals, then the graph is said to be *distance-transitive*.

In this setting, we will consider distance-regular graphs to be the combinatorial relaxation of being distance-transitive. For example, the Petersen graph, Higman-Sims graph, the Clebsch graph and most other named graphs are in fact distance-transitive. On the other hand, many Latin square graphs (and almost all strongly regular graphs) have trivial automorphism groups.

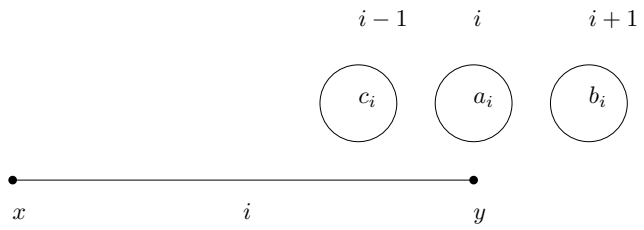


Figure 2.1: Vertices x, y at distance i and the neighbours of y partitioned by their distance to x.

Let X be a distance-transitive graph. Let $x, y \in V(X)$, let $i = d(x, y)$ and consider the neighbours of y and their distance to x. Any neighbour of y must be a distance $i-1$, i or $i+1$ from x. Let c_i, a_i, b_i be the number of neighbours of y at distance $i-1$, i or $i+1$ from x, respectively. Since X is distance-transitive, any pair of vertices (u, v) at distance i can be mapped to (x, y) under the action of the automorphism group and thus c_i, a_i, b_i are the numbers of neighbours of v at distance $i-1$, i or $i+1$ from u, respectively. See Figure 2.1.

If X is a graph of diameter d such that there exists constants c_i, a_i, b_i for $i = 0, \ldots, d$ such that any pair of vertices x, y at distance i has the property that c_i, a_i, b_i are the numbers of neighbours of y at distance $i-1$, i or $i+1$ from x, then X is said to be a *distance-regular graph*. Thus every distance-transitive graph is distance-regular.

Exercise 2.1. For a distance-regular graph, show that $\{a_i, b_i, c_i \mid i = 0, \ldots, d\}$ are determined by the *intersection array*, $\{b_0, \ldots, b_{d-1} : c_1, \ldots, c_d\}$.

A n-vertex graph X is *strongly regular* if X is neither complete or empty, every vertex has degree k, every pair of adjacent vertices have a common neighbours and every pair of non-adjacent vertices have c common neighbours. The tuple (n, k, a, c) are said to be the *parameter set* of X. A connected strongly regular graph is a distance-regular graph of diameter 2.

2.1 Examples

We will now give some examples of strongly regular and distance-regular graphs, in particular those constructions which yield pairs of graphs which may be very difficult to distinguish.

2.1.1 Paley graphs and Peisert graphs

Let $q \equiv 1 \mod 4$ be a prime power. Vertices of the *Paley graph* $P(q)$ are elements of $GF(q)$ and they are adjacent when their difference is a square. Paley graphs

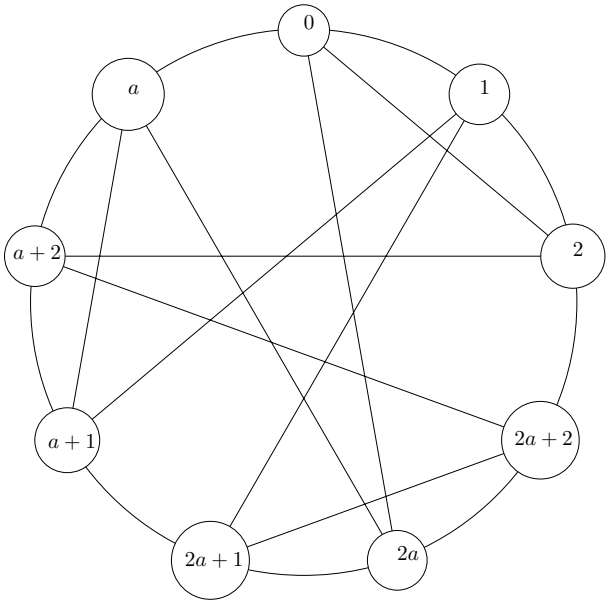


Figure 2.2: Paley graph over $\mathbb{F}_9 = \mathbb{F}_3(a)$ where $a^2 + 1 = 0$.

form one of two infinite families of graphs which are both self-complementary and arc-transitive. See Figure 2.2 for a small example.

The second infinite family of self-complementary and arc-transitive were given by Peisert in [20]. The vertices of the *Peisert graph* are the elements of $GF(q)$ where $q = p^r$ for $p \equiv 3 \mod 4$ a prime and r even. They are adjacent if their difference is in $\{a^j \mid j \equiv 0 \mod 4\}$ where a is a generator.

Theorem 2.2. *[20] Other than one exceptional graph on 23^2 vertices, the Paley and Peisert graphs are the only self-complementary and arc-transitive graphs.*

The Paley graph and Peisert graphs of order q are both strongly regular with parameters

$$\left(q, \frac{q-1}{2}, \frac{q-5}{4}, \frac{q-1}{4}\right).$$

Practically speaking, for q where the Peisert and Paley graphs are both define, this gives us a source of pair of graphs which have similar properties. The smallest q where both graphs are defined and not isomorphic to each other is $q = 49$.

2.1.2 Latin squares and orthogonal arrays

The following array is a *Latin square*:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{pmatrix}$$

For generally, a *Latin square* of order n is an $n \times n$ array where the symbols $1, \ldots, n$ occur exactly once in each row and in each columns.

Given a Latin square, the vertices of the *Latin square graph* are (i, j) for $i, j \in [n]$. They are adjacent if they are in the same row, same column, or share the same entry.

Exercise 2.3. Show that Latin square graphs are strongly regular and give their parameters.

Another way of recording the information in a Latin square is by putting it in an *orthogonal array*. The array above can be recorded as the following:

row	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4
col	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
entry	1	2	3	4	2	3	4	1	3	4	1	2	4	1	2	3

For example, the highlighted column indicates that the symbol in the row 2 and column 3 is 4. In general, for $t \leq k$, an orthogonal array of type (N, k, v, t), denoted $OA(N, k, v, t)$, is an $k \times N$ array whose entries are chosen from a v-element set X, such that in every subset of t rows of the array, every t-tuple of points of X is repeated the same number of times. The example above is an $OA(16, 3, 4, 2)$ and a Latin square of order n is an $OA(n^2, 3, n, 2)$. Note, here we have taken what many sources will call the transpose of the orthogonal array, mostly for ease of viewing the array.

Given an orthogonal array, we may construct the *orthogonal array graph* whoses the vertices are columns of the array and two columns are adjacent when they agree on some row.

Exercise 2.4. Show that the orthogonal array graph $OA(n^2, k, n, 2)$ are strongly regular and give their parameters.

2.1.3 Generalized quadrangles

An point-line incidence structure has points, lines and an incidence relation between points and lines. A *generalized quadrangle* is an incidence structure where each line *s*its on $s + 1$ points, there are $t + 1$ lines *t*hrough each point, such that

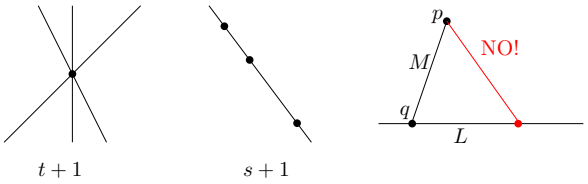


Figure 2.3: Picture of the definition of generalized quadrangle

for a point P and a non-incident line ℓ, there exist a unique point Q on ℓ which are collinear with P.

The point graph of a GQ has points for vertices and two points are adjacent if they are collinear (there is a line incident to both points). The parameters are

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

There are only a few families (s,t) where $GQ(s,t)$ is known to exist; they are

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

and their duals.

Open problem: The following values of (s,t) are feasible, but no generalized quadrangles are known to exist:

$$(q,q^2-q), (q,q^2-q-1).$$

In particular, does there exist $GQ(4,11)$ or $GQ(4,12)$? It is also unknown whether or not there exists two non-isomorphic $GQ(4,16)$.

For large q there are multiple constructions for $GQ(q,q^2)$, which could give non-isomorphic graphs, the smallest occurs with $q=5$. These are the strongly regular counterexamples to the Emms et al procedure of [10] given in [14].

2.1.4 Hadamard graphs

The Hadamard graph of order n is a distance-regular, antipodal and bipartite graphs constructed as follows. Let H be an $n \times n$ Hadamard matrix. The graph $\mathcal{H}$ of H has two vertices c^+ and c^- for each column of H and vertices r^+ and r^- for each column of H. For $1 \le i,j \le n$, the graph $\mathcal{H}$ has edges $r_i^+ c_j^+$ and $r_i^- c_j^-$ if $H(i,j) = 1$ and edges $r_i^+ c_j^-$ and $r_i^- c_j^+$ if $H(i,j) = -1$. The resulting graph has diameter 4 and $4n$ vertices.

Two Hadamard matrices H, H' are *Hadamard-equivalent* if H and obtained from H' by a series of the following operations:

- permuting rows;
- permuting column;
- multiplying rows by -1;
- multiplying columns by -1; and
- taking transposes.

Theorem 2.5 (McKay)**.** *The Hadamard graphs corresponding to H, H' are isomorphic if and only $H.H'$ are Hadamard-equivalent.*

For $n = 16, 20, 24$, there are multiple Hadamard graphs of order n, which are not distinguished by the spectrum the Emms et al matrix. For example, the Hadamard graph of order 4 (happens to be isomorphic to the 4-cube) has 26 elements in the 01 basis of its $W^{(2)}$.

Chapter 3

Matrix algebras

A *cellular algebra* is a subalgebra of the $n \times n$ matrices over $\mathbb{C}$ which is

(i) closed under $\circ$ (Schur/Hadamard/entry-wise multiplication);

(ii) closed under transposition and conjugation; and

(iii) contains J and I.

Here, and throughout the notes, $J_{m,n}$ denotes the $m \times n$ all ones matrix and I_n denotes the $n \times n$ identity matrix; we will omit the subscripts when the order of the matrices is clear from the context.

Given matrices $M_1, \ldots, M_\ell$, we can consider the *cellular closure*, denoted $\langle M_1, \ldots, M_\ell \rangle$, which is the smallest cellular algebra containing $M_1, \ldots, M_\ell$.

Example 3.1. For any graph X, we can consider the following algebra, $\mathcal{A} = \langle A(X) \rangle$. This is the *Bose-Mesner algebra* of X.

Exercise 3.2. A cellular algebra has a unique basis $\mathcal{A}$ of 01-matrices such that

(i) $\sum_{A \in \mathcal{A}} A = J$;

(ii) $\sum_{A \in \mathcal{A}' \subset \mathcal{A}} A = I$;

(iii) for each $A \in \mathcal{A}$, $A^T \in \mathcal{A}$;

(iv) for $A_i, A_j \in \mathcal{A}$ we have $A_i A_j = \sum_k p_{ij}^k A_k$ where p_{ij}^k are constants.

This basis is called the *Schur* basis, since its elements are Schur-idempotent (that is $A_i \circ A_i = A_i$) and pairwise orthogonal with respect to the Schur product.

If X is a distance-regular graph, then the unique 01-basis of $\mathcal{A} = \langle A(X) \rangle$ is $\{A_0 = I, A_1 = A(X), \ldots, A_d\}$, where

$$A_i(x, y) = \begin{cases} 1, & \text{if } d(x, y) = i; \\ 0, & \text{otherwise.} \end{cases}$$

We say that A_i is the i-th distance matrix. In this case, $\mathcal{A}$ has a second basis (not using 01-matrices); one can show that X has $d+1$ eigenspaces and the idempotent projectors onto the eigenspaces, $\{E_0 = \frac{1}{n}J, E_1, \ldots, E_d\}$, also forms a basis whose elements are idempotent and pairwise orthogonal with respect to the usual matrix product. Further, we can observe that $p_{1i}^{i-1} = b_{i-1}$ and so on; we can compute all other p_{ij}^k from the intersection array.

Exercise 3.3. Let $\mathcal{A}$ be a semi-simple, commutative cellular algebra of dimension $d+1$. Show that there are idempotent matrices $\{E_0, \ldots, E_d\}$ such that $ME_i = \lambda_{M,i}E_i$ for each $M \in \mathcal{A}$. Show that the change of basis between $\{A_0, \ldots, A_d\}$ and $\{E_0, \ldots, E_d\}$ depends only on $\{p_{ij}^k \mid 0 \leq i, j, k \leq d\}$. (This matrix is the P-eigenmatrix of the association scheme, which we will address later.)

3.1 Association schemes

A discussion of matrix algebras here would be remiss without discussing association schemes, which are a special case of cellular algebras. Suppose $n \times n$, symmetric 01-matrices $\{A_0, \ldots, A_d\}$ commute pairwise and share $d+1$ eigenspaces. Let $E_0, \ldots, E_d$ be the primitive idempotent projectors onto the shared eigenspaces of $\{A_i\}_{i=0}^d$. There exists constants p_{ij}^k, q_{ij}^k for $i, j, k \in \{0, \ldots, d\}$ such that following hold:

(i) $A_0 = I$ and $E_0 = \frac{1}{N}J$;

(ii) $\sum_{i=0}^{d} A_i = J$ and $\sum_{i=0}^{d} E_i = I$;

(iii) $A_i \circ A_j = \delta_{ij}A_i$ and $E_iE_j = \delta_{ij}E_i$; and

(iv) $A_iA_j = \sum_{k=0}^{d} p_{ij}^k A_k$ and $E_i \circ E_j = \sum_{k=0}^{d} q_{ij}^k E_k$. □

We say that $\mathcal{A} = \langle A_0, \ldots, A_d \rangle$ is an *(commutative) association scheme* and $\{A_0, \ldots, A_d\}$ are the *associate matrices* of $\mathcal{A}$. The constants p_{ij}^k are known as the *intersection numbers* of the scheme and the constants q_{ij}^k are known as the *Krein parameters* of the association scheme.

Since $\{E_i\}_{i=0}^d$ and $\{A_i\}_{i=0}^d$ are both bases of $\mathcal{A}$, there exists change of bases matrices between them. The *eigenmatrices* of the association scheme are $d+1 \times d+1$ matrices P and Q such that

$$A_j = \sum_{i=0}^{d} P_{ij}E_i \quad \text{and} \qquad E_j = \frac{1}{N}\sum_{i=0}^{d} Q_{ij}A_i. \tag{3.1}$$

Note that this implies that $\{P_{ij}\}_{i=0}^d$ are the eigenvalues of A_j.

Example 3.4. For some ordering, the P and Q matrices of the Hadamard graph of order n, from Section 2.1.4 are both equal to the following:

$$P = Q = \begin{pmatrix} 1 & n & 2n-2 & n & 1 \\ 1 & \sqrt{n} & 0 & -\sqrt{n} & -1 \\ 1 & 0 & -2 & 0 & 1 \\ 1 & -\sqrt{n} & 0 & \sqrt{n} & -1 \\ 1 & -n & 2n-2 & -n & 1 \end{pmatrix}.$$

Since $\mathcal{A}$ is an algebra and thus closed under multiplication and addition, there is a choice of the ordering of $A_1, \ldots, A_d$ such that A_i is a polynomial in A_1 for all i. If there is an ordering such that A_i is a polynomial in A_1 of degree i, for each $i = 0, \ldots, d$, we say the scheme is *P-polynomial*.

Exercise 3.5. Show that, up to reordering the associate matrices, the condition of being P-polynomial is equivalent to requiring that $p_{ij}^k = 0$ whenever the sum of two of $\{i, j, k\}$ is strictly smaller than the third element. This condition is also called *metric*.

In light of this, the terms P-polynomial and metric are often used exchangeably.

Exercise 3.6. The class metric schemes are exactly those where A_1 is the adjacency matrix of a distance-regular graph.

Similarly, we say a scheme is *Q-polynomial* if E_i is a polynomial under Schur multiplication in E_1 of degree i, for each $i = 0, \ldots, d$. A scheme is *cometric* if $q_{ij}^k = 0$ whenever the sum of two of $\{i, j, k\}$ is strictly smaller than the other element.

Exercise 3.7. A scheme is Q-polynomial if and only if it is cometric.

We have seen examples of association schemes arising from cellular algebras of the adjacency matrices of distance-regular graphs. There are many construction which are not metric.

Example 3.8. Let G be a group. Let $C_0, \ldots C_d$ be the conjugacy classes of G. Let A_i be as follows:

$$A_i(g, h) = \begin{cases} 1, & \text{if } gh^{-1} \in C_i; \\ 0, & \text{otherwise.} \end{cases}$$

Then $\{A_0, \ldots, A_d\}$ are the 01-basis of an association scheme. (The proof is left as an exercise.) This is called the *conjugacy class scheme* of G.

3.2 Cospectrality

It is well-known that if X, Y are distance-regular graphs with the same intersection array, then their adjacency matrices are cospectral. But also, the Laplacian

matrices, signless Laplacians, and Seidel matrices are also cospectral. In fact, for any $a_0, \ldots, a_d$ the matrices

$$M_X = \sum_{i=0}^{d} a_i A_i(X), \quad M_Y = \sum_{i=0}^{d} a_i A_i(Y)$$

are cospectral. We can now see that this is a direct consequence of Exercise 3.3.

This holds more generally to cellular algebras; but we need to "formalize" how two matrix algebras are "cospectral" in the above sense.

Exercise 3.9. The Wells graph (Armanios-Wells graph) is a distance-regular graph on 32 vertices with intersection array $\{5, 4, 1, 1; 1, 1, 4, 5\}$ has at least two cospectral mates. Can you construct them? Also, any cospectral mate will not be distance-regular; can you find a way to show that the cellular closures of their adjacency matrices are not weakly isomorphic?

3.3 Weak and strong isomorphism

Let W, W' be two cellular algebras with 01-bases $\mathcal{A}$ and $\mathcal{A}'$ respectively. We say that W and W' are *strongly isomorphic* if there exists an ordering of $\mathcal{A}'$ and a permutation matrix P such that

$$A_i = P^T A_i' P.$$

We say that W and W' are *weakly isomorphic* if there exists an ordering of $\mathcal{A}'$ such that there exist a map $\phi : \mathcal{A} \to \mathcal{A}'$ such that $\phi(A_i) = A_i'$ and preserves addition, multiplication, Schur multiplication, conjugation and transposition.

With this definition, W, W' will be weakly isomorphic if and only if there is an ordering of A' such that they have the same p_{ij}^ks.

To understand what this means, let us look at a distance-regular graph. The Bose-Mesner algebras of two distance-regular graphs X, Y are strongly isomorphic if and only if X and Y are isomorphic. This is clear since $A_1 = P^T A_1' P$. They are weakly isomorphic if and only if they have the same intersection array.

Lemma 3.10. *If W, W' are weakly isomorphic cellular algebras then*

$$\sum_{i=0}^{d} a_i A_i, \quad \sum_{i=0}^{d} a_i A_i'$$

are cospectral for any choice of $\{a_i\}$.

The proof is left as an exercise.

Chapter 4

Weisfeiler-Lehmann algorithm

For a graph X, a partition of its vertex set $V = V_1 \cup \cdots \cup V_d$ is said to be *equitable* if each vertex $v \in V_i$ has c_j neighbours in V_j for $0 \leq i, j \leq d$. If we use an equitable partition of X to partition $A(X)$ into block matrices, we would find that each block has a constant row sum. Every graph has an equitable partition; the partition into singletons is always equitable. If X is a regular graph, then the partition $\{V(X)\}$ is also equitable.

Example 4.1. Let x be a vertex of a distance-regular graph X of diameter d. Let X_i be the set of vertices at distance i from x. Show that the distance partition, $X_0 \cup \cdots \cup X_d$ is an equitable partition. Show that the converse is also true.

The Weisfeiler-Lehman algorithm is an algorithm which computes an equitable partition of an input graph. The 1-dimensional Weisfeiler-Lehman algorithm is also called *colour refinement*. In the k-dimensional WL method, we colour the k-tuples of vertex set. In the first step, we colour them with their isomorphism class. In subsequent steps, we colour each k-tuple with the multiset of colours of its neighbours (those which differ in exactly one position). We continue this way until an equitable partition is reached. For an example of this algorithm in action, see Figure 4.1.

Let X be a graph on n vertices. We will colour the elements of $V(X)^k$. For $S = (v_1, \ldots, v_k) \in V(X)^k$, we have that

$$S'(x) = \{(x, \ldots, v_k), \ldots, (v_1, \ldots, x)\}$$

for each $x \in V(G)$.

At time i of the k-dimensional WL algorithm, we will colour vertex v with $\ell_i(v)$. At the 0th iteration, $\ell_0(v)$ will be the isomorphism class of v, in the following way; $(x_1, \ldots, x_k)$ and $(v_1, \ldots, v_k)$ will have the same label if

1. $x_i = x_j$, then $v_i = v_j$; and
2. $x_i \sim x_j$, then $v_i \sim v_j$.

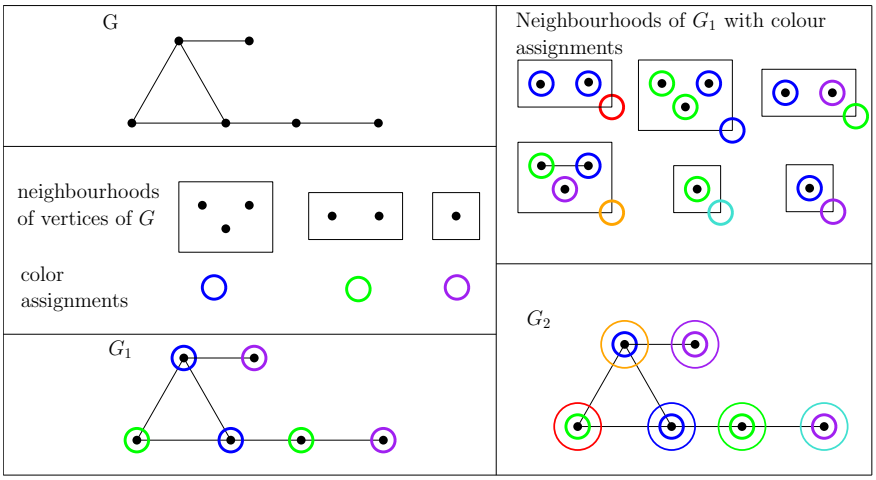


Figure 4.1: An example of colour refinement. At each step, we look at the vertex-coloured graphs induced by the vertices of the graph. We then colour the vertices with respect to the isomorphism class of the vertex-coloured graphs, and add this new colour to the tuple of colours already at the vertex. In G_2, the tuple of colour is shown by concentric circles. Note that the algorithm terminates at G_2 since an equitable partition (into singletons) has been reached.

For $i > 0$, we have for $S = (v_1, \ldots, v_k)$

$$\ell_i(S) = (\ell_{i-1}(S), \{\ell_{i-1}(T); T \in S'(x), \forall x \in V(X)\}).$$

Again, these ℓ_i will partition the k-tuples of the graph at every step. This will terminate and output some partition of the graph (and the labels).

We can use the Weisfeiler-Lehman algorithm as heuristic for graph isomorphism; if we run it on graph X and Y, if the set of labels and the partition is not able to distinguish X, Y, then we say that X, Y are not distinguished by the k-dimensional WL.

4.1 Cai-Fürer-Immerman graphs

A *separator* in a graph is $S \subset V$ such that the induced subgraph on $V - S$ has no connected component with $\geq |V|/2$ vertices. Following [8], we will construct a graph $X(G)$ from a given graph G, usually a low degree graph with linear size separators, and give a switching operation to obtain a second graph $X'(G)$. We will refer to the graph $X(G)$ as defined in this section as the *Cai-Fürer-Immerman graph* of G.

For each positive integer d, we define a graph X_d as follows: X_d has vertex set V_d and edge set E_d. The vertex set consists of 3 disjoint sets, A_d, B_d and M_d such that

A_d and B_d each contain d elements, indexed by $\{1, \ldots, d\}$ and M_d contains one element indexed by each even subset of $\{1, \ldots, d\}$. Vertices $a_i \in A_d$ and $m_S \in M_d$ are adjacent if $i \in S$ and vertices $b_j \in B_d$ and $m_S \in M_d$ are adjacent if $j \notin S$.

To obtain the Cai-Fürer-Immerman graph of G, we replace each vertex v of G with a graph $X(v)$ where $X(v) = X_d$ and d is the valency of v. In $X(v)$ we associate one pair of vertices $\{a_i, b_i\}$ with each neighbour w of v and write $a(v, w) = a_i$ and $b(v, w) = b_i$. For each edge v, w in G, we add the edges $\{a(v, w), a(w, v)\}$ and $\{b(v, w), b(w, v)\}$.

A *twist* of $X(G)$, which we will denote $X(G)$, is obtained by choosing an edge of G and replacing the edges $\{a(v, w), a(w, v)\}$ and $\{b(v, w), b(w, v)\}$ in $X(G)$ with $\{a(v, w), b(w, v)\}$ and $\{b(v, w), a(w, v)\}$.

If we take the complete graph on 4 vertices, then $X(K_4)$ has 40 vertices and is regular with valency 4, see Figure 4.2. Let $\widetilde{X}(K_4)$ be the twist of $X(K_4)$. It can be verified that $X(K_4)$ and $\widetilde{X}(K_4)$ are not isomorphic but $S^+\left(U\left(X(K_4)\right)^3\right)$ and $S^+\left(U\left(\widetilde{X}(K_4)\right)^3\right)$ have the same spectrum.

4.2 Connections to Weisfeiler-Lehman

There are several other ways of understanding graph which are not distinguished by the k-dimensional Weisfeiler-Lehman algorithm. It is an active area of research. We will give some of them for some context.

4.2.1 Homomorphism counts

Let $\hom(F, G)$ denote the number of homomorphism from a graph F to a graph G. There is a well-known theorem of Lovász which determines isomorphism using homomorphism counts.

Theorem 4.2. *[17] Let G, H be graphs. $\hom(F, G) = \hom(F, H)$ for all graphs F if and only if G is isomorphic to H.*

Note that if $\hom(F, G) = \hom(F, H)$ for all cycles F, then G and H are cospectral.

A more recent result of Dell, Grohe, Rattan in 2018 relates homomorphism counts to the Weisfeiler-Lehman algorithm.

Theorem 4.3. *[9] If $\hom(F, G) = \hom(F, H)$ for all trees F, then colour refinement does not distinguish G, H.*

Theorem 4.4. *[9] If $\hom(F, G) = \hom(F, H)$ for all graphs F of treewidth at most k, then k-dimensional WL does not distinguish G, H.*

In the same vein, Mančinska and Roberson [18] show that quantum isomorphism is equivalent to equality of homomorphism counts from planar graphs.

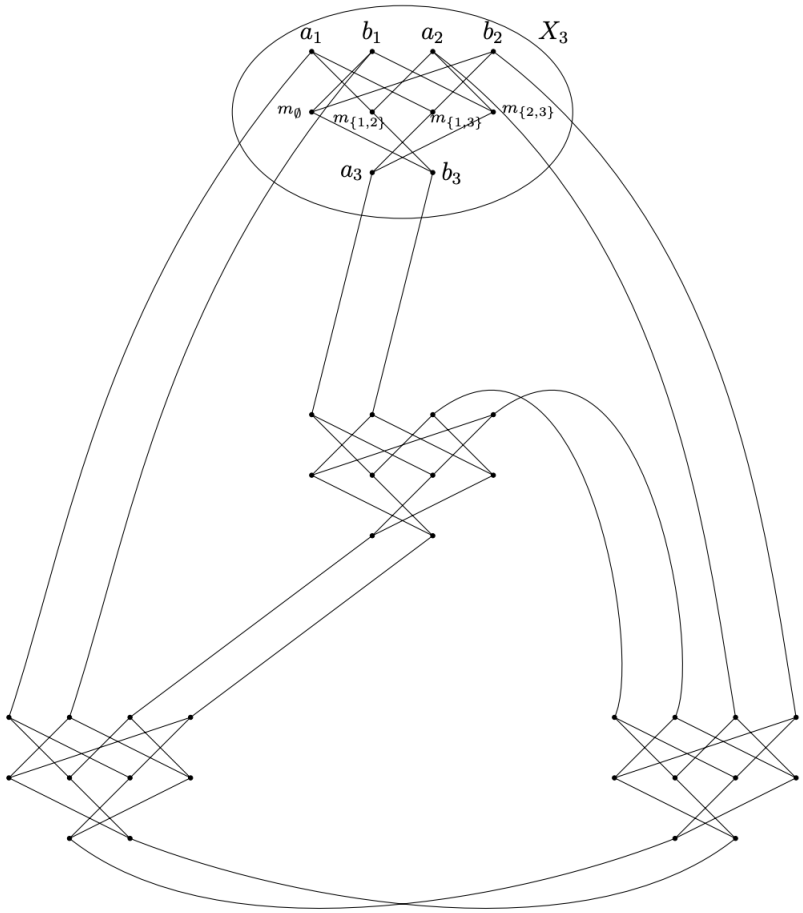


Figure 4.2: The Cai-Fürer-Immerman graph of K_4.

4.2.2 Symmetric powers

The k-th symmetric power of a graph X is a graph such that the vertices are the k-subsets of X and are adjacent when the symmetric difference is an edge. It was proposed by Audenaert et al that the spectrum of the k-th symmetric power will distinguish graphs.

Theorem 4.5. *[1] If the $2k$-dimensional WL does not distinguish X, Y, then the k-th symmetric powers of X, Y will be cospectral.*

There is a construction of Cai, Fürer and Immerman 1992 which gives, for each k, pairs of graphs which are not distinguished by k-dim WL. However, it is an open problem to find a pair of strongly regular graphs which are not distinguished by the spectra of their k-th symmetric power, where $k \geq 3$.

Chapter 5

Extensions of matrix algebras

Let W be a cellular algebra. For us, we usually take W to be the cellular closure of the adjacency matrix of graph X. We will now define the kth extension of W; this extension will contain $W^{\otimes k}$, but also many other matrices. For most of this, we will follow the treatment in [21].

Consider Δ_I, a matrix indexed by the elements of $V(X)^k$. It is a diagonal matrix where $\Delta_I(x, y) = 1$ whenever $x, y = (u, u, \ldots, u)$. The *k-extension* of W is

$$W^{(k)} = \langle W^{\otimes k}, \Delta_I \rangle.$$

It is easy to see that this is bigger than just the tensor product. Colloquially speaking, we can get a sense of how much bigger it is by having a look at the cells of the cellular algebra.

Let $\mathcal{A}$ be the 01-basis of W. Let T be an array of basis matrices; $T = \{A_{i,j}\}_{1 \le i,j \le k}$. A k-tuple $x \in V^k$ is type T if

$$A_{i,j}(x_i, x_j) = 1$$

for all i, j. Let Δ_T be the identity on the k-tuples of type T. Then, $\Delta_T \in W^{(k)}$. In particular, the cell partition of $W^{(k)}$ must be a refinement of the partition into types.

5.1 Equivalence of extensions

Let X, Y be graphs with adjacency matrices A, A', respectively, and cellular closure W, W' respectively. We will say that X, Y are equivalent if there is a weak isomorphism ϕ such that $\phi(A) = A'$.

For k-extension, we say that X, Y are k-equivalent if they are equivalent (via ϕ) and there exist a weak isomorphism $\widehat{\phi} : W^{(k)} \to W'^{(k)}$ such the restriction of $\widehat{\phi}$ to $W^{\otimes k}$ is ϕ^k and $\widehat{\phi}(\Delta_I) = \Delta_I$. We will say $\widehat{\phi}$ is a k-weak isomorphism.

Theorem 5.1. *[11] If the k-dimensional Weisfeiler-Lehman does not distinguish X and Y, then X, Y are $\lfloor \frac{k}{3} \rfloor$-equivalent. If X, Y are k-equivalent then the k-dimensional WL does not distinguish them.*

We want to relate this to our discussion of graph invariants in Sections 1.3.1 and 1.3.2.

Theorem 5.2. *[1, 6] If X, Y are k-equivalent and $\widehat{\phi}$ is a k-weak isomorphism from $W^{(k)}$ to $W'^{(k)}$, then $\widehat{\phi}$ takes the kth symmetric power of X to the kth symmetric power of Y.*

Theorem 5.3. *[21] If X, Y are k-equivalent and $\widehat{\phi}$ is a k-weak isomorphism from $W^{(k)}$ to $W'^{(k)}$, then $\widehat{\phi}(U_X^k) = U_y^k$ and $\widehat{\phi}(S^+(U_X^k)) = S^+(U_y^k)$. In particular, if X, Y are k-equivalent, then $S^+(U(X)^k)$ and $S^+(U(Y)^k)$ are cospectral.*

In other words, Smith shows that the procedure of Emms et al cannot distinguish a pair of graphs which are 3-equivalent. It is however still very difficult to find such strongly regular pairs.

Open problem: Find a pair of strongly regular graphs which are 3-equivalent?

5.2 Triply regular graphs

In order to answer the question posed at the end of the previous section, a promising source is to look at triply regular graphs and triply regular association schemes. We will again motivate our definition of triply regular using group actions. The automorphism group of a graph also acts on the 3-tuples of vertices. What can we say about the orbits of this action? Suppose we have (u, v, w) and (x, y, z). If they are in the same orbit then

$$d(x, y) = d(u, v), \quad d(y, z) = d(v, w), \quad d(x, z) = d(u, w).$$

We can partition the triples of vertices by the pairwise distances; the orbit partition must be a refinement of this partition. The graph is said to be *triply transitive* if this is the orbit partition. Triply regular is the combinatorial relaxation of triply transitive; a graph is *triply regular* if the constants p_{ijk}^{rst} exists.

Examples of triply regular graphs include the hypercube, line graph of $K_{n,n}$, half cube, Hadamard graphs of diameter 4, the double cover of the Higman-Sims and any strongly regular graph where $q_{ii}^i = 0$ for $i \in \{1, 2\}$. The point graph of $GQ(q, q^2)$ is also triply regular.

Open problem: If X, Y are triply regular graphs with the same triple intersection numbers then they are 2-equivalent? Even 3-equivalent?

This appears to be a difficult problem because it appears that the k-extensions are rather difficult matrix algebras to work with. One possible of attack is to relate them to matrix algebras which are more tractable because they admit a more combinatorial definition. We will give an overview of some of these matrix algebras in the next section.

5.3 Other matrix algebras

5.3.1 Terwilliger algebras

We will now look at the Terwilliger algebra, which was defined by Terwilliger in [23]. We consider an association scheme $\mathcal{A}$ with associate matrices $A_0, \ldots, A_d$. Throughout this section, let $x \in V$ be a fixed vertex of $\mathcal{A}$. For $i = 0, \ldots, d$, we will define the diagonal matrix $E_i^*(x)$ as follows:

$$E_i^*(x)_{y,y} = (A_i)_{x,y}.$$

We call $E_i^*(x)$ the *ith dual idempotent with respect of x*. Similarly, for $i = 0, \ldots, d$, we will consider diagonal matrices $A_i^*(x)$ with entries as follows:

$$A_i^*(x)_{y,y} = n(E_i)_{x,y},$$

where $n = |V|$. We call A_i^* the *ith dual distance matrix with respect to x*. When the context is clear, we will write E_i^* for $E_i^*(x)$ and A_i^* for $A_i^*(x)$.

Note, that if A_1 is the adjacency matrix of a distance regular graph, E_i^* is the diagonal characteristic matrix for the set of vertices at distance i from x, also known as the *ith neighbourhood of x*. We also have that

$$A_i^* A_j^* = \sum_{k=0}^{d} q_{ij}^k A_k^*.$$

The matrices $A_0^*, \ldots, A_d^*$ form a basis for a subspace $\mathcal{A}^*$ of $n \times n$ complex (diagonal) matrices. In $\mathcal{A}^*$, the primitive idempotents are $E_0^*, \ldots, E_d^*$. Recalling the eigenmatrices of a scheme, we can give the change of basis in $\mathcal{A}^*$ as follows:

$$E_j^* = \frac{1}{N} \sum_{i=0}^{d} P_{ij} A_i^* \text{ and } A_j^* = \sum_{i=0}^{d} Q_{ij} E_i^*.$$

The *Terwilliger algebra* $T(x)$ is the subalgebra of $n \times n$ complex matrices generated by $\mathcal{A}$ and $\mathcal{A}^*$.

5.3.2 Jaeger algebras

We consider a symmetric association scheme $\mathcal{A} = \langle A_0, \ldots, A_d \rangle$. We will consider the following endomorphism of $n \times n$ matrices

$$X_i(M) = A_i M, \quad Y_i(M) = M A_i, \quad \Delta_i(M) = A_i \circ M.$$

These encode left multiplication, right multiplication and Schur multiplication by elements of $\mathcal{A}$, respectively. The *Jaeger algebra* $\mathcal{J}_2$ is generated by

$$\{X_0, \ldots, X_d\}.$$

The *Jaeger algebra* $\mathcal{J}_3$ is generated by

$$\{X_0, \ldots, X_d\} \cup \{\Delta_0, \ldots, \Delta_d\}.$$

The *Jaeger algebra* $\mathcal{J}_4$ is generated by

$$\{X_0, \ldots, X_d\} \cup \{\Delta_0, \ldots, \Delta_d\} \cup \{Y_0, \ldots, Y_d\}.$$

These algebra were defined by Jaeger for the purposes of studying spin models and representations of the braid group, see [16].

5.3.3 Connections between matrix algebras

These matrix algebras closely related; one can show that a restriction of $\mathcal{J}_3$ to some of its modules is isomorphic to the Terwilliger algebra.

Theorem 5.4. *[23] The set of simple modules of $\mathcal{J}_3$ is the disjoint union of the simple modules for each of the Terwilliger algebras $T(x)$ for $x \in V(X)$.*

Together with a result of Munemasa from [19], this implies the following.

Corollary 5.5. *If X, Y are triply regular with the same triple intersection numbers, then their $\mathcal{J}_3$ algebras are weakly isomorphic.*

This leaves us with many open problems, especially relating k-equivalence and weak isomorphism between Terwilliger and Jaeger algebras.

Open problem: Is weak equivalence of $\mathcal{J}_3$ of graph X and Y equivalent to 2-equivalence of X and Y?

Open problem: Is weak equivalence of $\mathcal{J}_4$ of graph X and Y equivalent to 3-equivalence of X and Y?

References

[1] A. Alzaga, R. Iglesias, and R. Pignol. Spectra of symmetric powers of graphs and the Weisfeiler-Lehman refinements. *J. Combin. Theory Ser. B*, 100(6):671–682, 2010.

[2] K. Audenaert, C. Godsil, G. Royle, and T. Rudolph. Symmetric squares of graphs. *J. Combin. Theory Ser. B*, 97(1):74–90, 2007.

[3] L. Babai. Laszlo Babai's Home Page. `http://people.cs.uchicago.edu/~laci/`. [Accessed 23-08-2023].

[4] L. Babai. Graph isomorphism in quasipolynomial time, 2016.

[5] L. Babai. Graph isomorphism in quasipolynomial time (extended abstract). In *Proceedings of the 48th annual ACM SIGACT symposium on theory of computing, STOC '16, Cambridge, MA, USA, June 19–21, 2016*, pages 684–697. New York, NY: Association for Computing Machinery (ACM), 2016.

[6] A. R. Barghi and I. Ponomarenko. Non-isomorphic graphs with cospectral symmetric powers. *Electron. J. Comb.*, 16(1):research paper r120, 14, 2009.

[7] A. E. Brouwer and E. Spence. Cospectral graphs on 12 vertices. *Electron. J. Comb.*, 16(1):research paper n20, 3, 2009.

[8] J.-Y. Cai, M. Furer, and N. Immerman. An optimal lower bound on the number of variables for graph identification. In *30th Annual Symposium on Foundations of Computer Science*, pages 612–617, Washington, DC, USA, 1989. IEEE Computer Society.

[9] H. Dell, M. Grohe, and G. Rattan. Lovász meets Weisfeiler and Leman. In *45th international colloquium on automata, languages, and programming. ICALP 2018, Prague, Czech Republic, July 9–13, 2018. Proceedings*, page 14. Wadern: Schloss Dagstuhl – Leibniz Zentrum für Informatik, 2018. Id/No 40.

[10] D. Emms, E. R. Hancock, S. Severini, and R. C. Wilson. A matrix representation of graphs and its spectrum as a graph invariant. *Electron. J. Comb.*, 13(1):research paper r34, 14, 2006.

[11] S. Evdokimov and I. Ponomarenko. On highly closed cellular algebras and highly closed isomorphisms. *Electron. J. Comb.*, 6(1):research paper r18, 31, 1999.

[12] J. K. Gamble, M. Friesen, D. Zhou, R. Joynt, and S. N. Coppersmith. Two-particle quantum walks applied to the graph isomorphism problem. *Phys. Rev. A*, 81:052313, May 2010.

[13] C. Godsil. *Algebraic combinatorics*. Chapman and Hall Mathematics Series. Chapman & Hall, New York, 1993.

[14] C. Godsil, K. Guo, and T. G. J. Myklebust. Quantum walks on generalized quadrangles. *Electron. J. Combin.*, 24(4):Paper No. 4.16, 6, 2017.

[15] C. Godsil and G. Royle. *Algebraic graph theory*, volume 207 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2001.

[16] F. Jaeger. On spin models, triply regular association schemes, and duality. *J. Algebraic Combin.*, 4(2):103–144, 1995.

[17] L. Lovász. Operations with structures. *Acta Math. Acad. Sci. Hungar.*, 18:321–328, 1967.

[18] L. Mančinska and D. E. Roberson. Quantum isomorphism is equivalent to equality of homomorphism counts from planar graphs. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, pages 661–672. IEEE Computer Soc., Los Alamitos, CA, [2020] ©2020.

[19] A. Munemasa. An application of terwilliger's algebra, March 1993. `http://www.math.is.tohoku.ac.jp/~munemasa/unpublished.html`.

[20] W. Peisert. All self-complementary symmetric graphs. *J. Algebra*, 240(1):209–229, 2001.

[21] J. Smith. *Algebraic aspects of Multi-Particle Quantum Walks*. Phd thesis, University of Waterloo, Waterloo, Ontario, Canada, June 2012.
Available at `http://hdl.handle.net/10012/7138`.

[22] D. A. Spielman. Faster isomorphism testing of strongly regular graphs. In *Proceedings of the 28th annual ACM symposium on the theory of computing, STOC '96. Philadelphia, PA, USA, May 22–24, 1996*, pages 576–584. New York, NY: ACM, 1996.

[23] P. Terwilliger. The subconstituent algebra of an association scheme. I. *J. Algebraic Combin.*, 1(4):363–388, 1992.

Part II

An Introduction to Code-Based Cryptography

Anna-Lena Horlemann

School of Computer Science,
University of St. Gallen,
Torstrasse 25,
9000 St. Gallen,
Switzerland

email: anna-lena.horlemann@unisg.ch

Contents

1 Introduction **39**

2 Preliminaries **41**
2.1 What is cryptography? 41
2.2 Asymmetric and public key cryptography 42
2.3 Attacks and cryptanalysis 44
2.4 The issue with quantum computers 45
2.5 Error-correcting codes 46

3 Code-Based Cryptosystems **51**
3.1 General setup for public key encryption 51
3.2 Overview of some variants 54
3.3 Variants in the Hamming metric 55
3.4 Variants in the rank metric 60
3.5 Other metrics, other alphabets, other cryptosystems (outlook) . . . 65

Bibliography **67**

Chapter 1

Introduction

Most of our currently used cryptographic systems rely on either the integer factorization problem or the discrete logarithm problem over an elliptic curve or a finite field. We say that a problem is "hard" for cryptographic purposes if there is no polynomial time algorithm (known) to solve these problems. For the three problems above (integer factorization and two versions of the discrete logarithm problem) this is true for conventional computers. On quantum computers, however, there is a known algorithm that solves these problems in polynomial time. This algorithm is known as *Shor's algorithm*[1]; it was originally formulated for integer factorization, but can be adapted to solve discrete logarithm type problems, as well. This means that the classical systems are not secure and new algorithms are necessary for the future, to ensure secure communication in the time of quantum computers.

With the recent progression of quantum computer development, the cryptographic research community has therefore been actively looking for cryptographic systems that can withstand attacks from quantum computers. This realm of study is referred to as *post-quantum cryptography*.

At the time of writing there are five main streams of post-quantum cryptography:

- code-based cryptography
- lattice-based cryptography
- multivariate cryptography
- hash-based cryptography
- supersingular elliptic curve isogeny cryptography

In 2016, the National Institute of Standards and Technology (NIST) initiated a standardization procedure for post-quantum cryptosystems. These cryptosys-

[1] `https://en.wikipedia.org/wiki/Shor's_algorithm`

tems typically derive their foundation from NP-complete problems, primarily for two reasons: NP-complete problems are at least as complex as the most challenging problems within NP, while their solutions can be efficiently verified. In the current NIST standardization project almost all submissions for public-key encryption in the third round are either code or lattice based. In this lecture we will focus on the first, *code-based cryptography*.

Code-based cryptography relies on the fundamental premise that decoding in a randomly generated linear code constitutes an NP-complete problem, as demonstrated by Berlekamp, McEliece, and Van Tilborg in 1978 [2]. In that same year, McEliece introduced a cryptosystem [17] where a strategically chosen code with an inherent algebraic structure and an efficient decoding algorithm is employed. This code is then masked to appear as a random linear code. The process involves encoding a message into a codeword and encrypting it by introducing an error to the message. With knowledge of the code's algebraic structure, the original message can be retrieved. However, an adversary is confronted with the challenge of decoding an erroneous codeword in a randomly generated linear code.

These lecture notes provide an introduction to code-based cryptography, delving into the mathematical underpinnings of such systems and the challenges associated with designing secure yet practical schemes. We explore the main ideas of code-based public key encryption schemes and the strategies employed to potentially breach these cryptographic systems. Naturally, these lecture notes cannot cover all aspects of code-based cryptography and we refer the interested reader to the nice survey [26] and the references therein for more information.

Chapter 2

Preliminaries

2.1 What is cryptography?

We want to communicate (or store) data in a secure manner, such that no eavesdropper can recover the sent information. This is done with the help of *cryptography*, by *encrypting* the data before sending it, and then *decrypting* it at the receiver's side. An encryption is secure if an eavesdropper, who can intercept the encrypted message, cannot recover the original message from it. Cryptographic security depends on the computing power of the adversary. We call something *cryptographically secure* for a prescribed time t, if the best (known) way of breaking the cryptographic instance needs more than time t, on a given computer (possibly a personal computer or a large scale mega-computer – depending on the application). Moreover, if information is to remain confidential for many years, we have to take the current and future development of computers into account. In particular, new protocols for long-term use should be secure against attacks on a quantum computer.

The original idea of cryptography is to encrypt data such that only the intended receiver can recover the actual message, whereas eavesdroppers cannot. We distinguish two classes of such algorithms: *symmetric* and *asymmetric* encryption schemes. In the former the sender and receiver both use the *same key* in the en- and decryption procedure[1], whereas in the latter the sender encrypts with one key (called the *public key*), and the receiver decrypts with a different key (called the *private/secret key*)[2].

A different, but related, type of algorithm is a digital signature. Such signatures can be used to verify at the receiver's side if the received message was indeed from the sender, or if it was tampered with (or even replaced) by the adversary.

[1] This can be depicted by a normal lock for which both parties hold a key.

[2] This can be depicted with a padlock, where the public key is the actual padlock itself and the secret key is the key to the padlock, see Section 2.2.

Since symmetric cryptosystems are (so far) not drastically effected by quantum computers, we will focus on asymmetric cryptosystems and digital signatures.

2.2 Asymmetric and public key cryptography

Public key cryptography is a major subfield of asymmetric cryptography. In a public key encryption system some sender, say Alice, wants to send an encrypted message to a receiver, say Bob. The cryptosystem is asymmetric, in the sense that Bob publishes a *public key* and secretly stores a *private key*, such that Alice (and everyone else) can use the public key to encrypt her message, and only Bob can decrypt the message (with his private key). For the system to be secure, an attacker should not be able to decrypt the encrypted message without the knowledge of the private key. This is done by using some "hard" mathematical problem.

Digital signatures can also be seen as asymmetric cryptosystems, in particular as reversed versions of public key cryptosystems. Here, Bob uses his private key to sign a message, and Alice (or anyone) can verify his signature with the aid of the public key.

Remark 2.1. Asymmetric cryptography can be depicted with a padlock. If Alice wants to send a message in a treasure chest to Bob, he can send her an open padlock (the public key) that only he has a key to (the private key). She can use the lock on the chest and send it to Bob. He can use his private key to unlock the chest.

Now, of course, the question remains, how to construct such padlocks mathematically.

Mathematical idea (one-way function): Find "something" that is easy to compute, but where it is very (very very) hard to compute the inverse.
$\Longrightarrow$ Then the outcome of the computation is the public key and the preimage is the private key.

As an example we will look at the RSA cryptosystem, which is based on the integer factorization problem. Here we use the fact that it is easy to compute the product of two integers, but it is not easy to find the factors of a given product (of very large numbers). The algorithm is described in Algorithm 1.

Algorithm 1 RSA Algorithm

- Choose two large prime numbers p and q.
- Compute the product $n = p \cdot q$.
- Pick any number e that is relatively prime to $(p-1)(q-1)$.
- Compute $d \equiv e^{-1} \mod (p-1)(q-1)$.
- **Private key:** The number d.
- **Public key:** The numbers n and e.
- **Encryption:** Represent the message as a number m. It is encrypted as

$$c \equiv m^e \mod n.$$

- **Decryption:** Compute

$$m \equiv c^d \mod n.$$

Let's have a look at why the underlying hard problem of RSA is the integer factorization problem – if an attacker manages to factor n into p and q, they can recover the private key and hence decrypt any ciphertext. In fact, almost all currently used asymmetric cryptosystems use one of the following three mathematical problems:

- integer factorization
- discrete logarithm problem
- elliptic curve discrete logarithm problem

As we will see, these three problems are not hard to solve on quantum computers and hence need to be replaced by other mathematical problems in quantum secure cryptosystems.

Exercise 2.2. Create an RSA signature scheme, by first using the private key for signing, and the public key for validating a signature. Is Alice's signature always the same or does it depend on the message?

Exercise 2.3. Assume that Alice wants to send a secret message to Bob in a chest/box with Bob's open padlock. If the eavesdropper Eve is the mail woman and transports the open padlock of Bob, as well as the locked chest of Alice – how can she trick Alice and Bob and read the message?

Exercise 2.4. Assume that Alice and Bob still want to communicate in a secure manner by sending a message in a chest/box to each other. Assume furthermore

that there is no way of providing (reliably) a public open padlock. However, Alice and Bob both have a padlock with a key. How can Alice still send a message secretly to Bob?

2.3 Attacks and cryptanalysis

To be pessimistic up front: any cryptosystem (no matter if digital or analog) can be broken. Full stop. However, when we speak of *security* we assume that it is extremely unlikely that an attacker can break a system. For this we *measure* the security of a system in the number of operations an attacker needs (worst case or on average) to break a system.[3] If the number of operations needed for an attack supersedes the number of atoms in the Universe (ca. 2^{80}) we consider this a first level of security (for non-sensitive information). In most real-world applications, however, we rather use systems with a security level of 256 or more bits (i.e., 2^{256} necessary operations for an attack).

We distinguish two main types of theoretical attacks on an encryption scheme (and similarly on a signature scheme):

- **Message recovery attacks:** The attacker can directly decrypt a ciphertext but does not recover the used (private) key.
- **Key recovery attacks:** The attacker recovers the (private) key and can thus also decrypt any ciphertext.

Both are equally important[4] and need to be considered when analyzing a cryptosystem. However, they often use different techniques and tools and are therefore considered separately.

A special type of – and commonly used – key recovery attack is a *distinguisher attack*. In this case the attacker uses publicly known (mathematical) properties of the private key (that distinguishes it from a random element of the ambient set) to recover the private key from the public key. We will look at some distinguisher attacks in the following chapter.

Exercise 2.5. There are several security considerations for the setup of RSA. Here are some examples:

1. You should always choose two distinct primes $p \neq q$. How can Eve attack RSA if $p = q$?

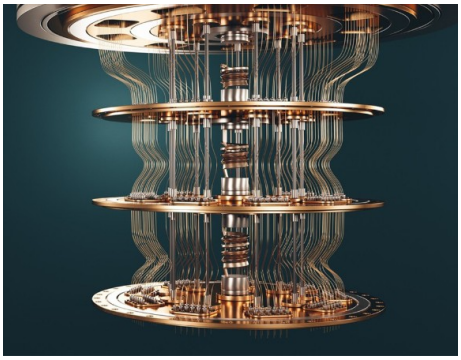
[3] This is more general than considering the actual time a specific computer would need to break the system.

[4] To be precise, the latter usually has a more severe impact on the long-term use of the cryptosystem, but both types need to be infeasible for a system to be secure.

2. When sending the same message to several parties, you should always use different moduli n. How can Eve recover the message m if she intercepted two ciphertexts $c_1 = m^{e_1} \mod n$ and $c_2 = m^{e_2} \mod n$?

2.4 The issue with quantum computers

Quantum computers are one of the most exciting and challenging research objects in today's communication and computation research. Due to their different functionality, they promise big advantages in many computational applications such as simulations in biotechnology, chemistry, pharmaceutical research, etc., and optimization tasks in finance, energy management, or logistics.



On the other hand, quantum computers pose one of the biggest threats in cybersecurity, in particular to public key cryptosystems and digital signatures. Currently the most commonly used public key cryptosystems are RSA and Diffie-Hellman (in classical and elliptic curve variants). All these methods are insecure if quantum computers capable of running Shor's algorithm[5] become available, which is likely to happen in the next few decades. The importance of this threat is further stressed by the NSA's statement in 2015 to transition to post-quantum secure algorithms for their Suite B family of cryptographic algorithms and the National Institute of Standards and Technology's (NIST) initiation of the international Post-Quantum Cryptography Standardization project[6] in 2016.

Remark 2.6. It is not true that quantum computers will speed up every type of algorithm. In fact, there are only very few algorithms known where the quantum technology will lead to a significant speed up.[7]

The expected time frame for quantum computers to be powerful enough to break these implementations is between 10 to 30 years, however there are two major reasons why we should transition to post-quantum cryptography already now:

[5] https://en.wikipedia.org/wiki/Shor's_algorithm

[6] https://csrc.nist.gov/Projects/post-quantum-cryptography/post-quantum-cryptography-standardization

[7] https://en.wikipedia.org/wiki/Quantum_algorithm

1. Adversaries can store captured encrypted data for many years and decrypt it when the development of quantum computers enables the adversary to do so. This is a particular threat for highly sensitive data that needs to be secured over a long period of time, such as diplomatic data or trade secrets.
2. Communication systems in aircrafts, trains or ships (among others) are often in operation for more than 30 years and are hardly ever updated due to operational reasons. It is thus of paramount importance that communication algorithms on such systems are quickly changed to post-quantum secure systems. Otherwise these systems will be vulnerable and become accessible to adversaries.

Many proposals for post-quantum secure encryption and digital signatures have been made in the past, many of which have been broken by now. Currently, the National Institute of Standards and Technology (NIST) is considering a number of proposals for post-quantum secure encryption and signatures to be made US (and most likely also international) standards. The goal is to find cryptosystems of comparable size and computational complexity to RSA (or Diffie-Hellman), but that are secure against attacks run on quantum computers.

Remark 2.7. Since Shor's algorithm is currently the only known algorithm that leads to a speed up from exponential to polynomial run time on quantum computers, post-quantum cryptography can also be reformulated as the Shor-resistant cryptography.

As already mentioned in the introduction there are currently five main streams of *post-quantum cryptography*, based on different hard mathematical problems:

- code-based cryptography
- lattice-based cryptography
- multivariate cryptography
- hash-based cryptography
- supersingular elliptic curve isogeny cryptography

We will focus on *code-based cryptography*, which is based on tools from *coding theory* or *error-correcting codes*, which we will introduce in the following section.

2.5 Error-correcting codes

Since code-based cryptography uses a lot of *coding theory*, we start by briefly explaining *error correcting codes*. This will later on be used to explain the basic ideas

of code-based cryptography.

Error correcting codes are classically used for communication over a noisy channel. One adds redundancy to the data to be sent, so that the noise (or the errors) can be filtered out on the receiver's side. Depending on the channel, the noise type varies and different codes have to be constructed for the different channel types. The differences can lie in the choice of the underlying alphabet, as well as the choice of the coding metric that is used. We will introduce the alphabets and metrics that will be used in these lecture notes in the following. For more details we refer the reader to the books [13, 15, 16].

Denote by $\mathbb{F}_q$ the finite field with q elements where q is a prime power.

Definition 2.8. A classical *block code* is a set of vectors of a fixed length n over some finite field $\mathbb{F}_q$. A block code in $\mathbb{F}_q^n$ is called *linear*, if it is a linear subspace of $\mathbb{F}_q^n$.

A linear code $C \subseteq \mathbb{F}_q^n$ of dimension k can be represented by a *generator matrix* $G \in \mathbb{F}_q^{k\times n}$ such that $C = \langle G\rangle$ – where $\langle G\rangle$ denotes the row space of a matrix G – or a *parity check matrix* $H \in \mathbb{F}_q^{(n-k)\times n}$ such that C is the kernel of H.

Even though these are the most studied families, error-correcting codes do not have to consist of vectors. For example, in *linear network coding* codewords are matrices over $\mathbb{F}_q$, or they are the row spaces of those matrices. In the former case, the codewords are still elements of the (matrix) vector space $\mathbb{F}_q^{m\times n}$, for some $m, n \in \mathbb{N}$, and hence linearity of such matrix codes can be defined in the usual sense. For our lecture notes the notion above (where codewords are vectors) is sufficient.

Various metrics can be used for different coding theoretic applications. In this lecture we will consider the Hamming and the rank metric. We will explain these metrics in the following.

Definition 2.9.

- The *Hamming distance* d_H on $\mathbb{F}_q^n$ is defined as

$$d_H((u_1,\dots,u_n),(v_1,\dots,v_n)) := |\{i \mid u_i \neq v_i\}|$$

  for any $(u_1,\dots,u_n),(v_1,\dots,v_n) \in \mathbb{F}_q^n$. Note that this metric can be defined for vectors over any underlying alphabet.

- A *rank metric code* is a subset of the matrix space $\mathbb{F}_q^{m\times n}$. The *rank distance* d_R is defined as the rank of the difference of the corresponding matrices over $\mathbb{F}_q$, i.e., for $A, B \in \mathbb{F}_q^{m\times n}$,

$$d_R(A,B) := \operatorname{rank}(A-B).$$

  Note that these codes can also be represented in $\mathbb{F}_{q^m}^n$, via a vector space isomorphism $\mathbb{F}_q^m \cong \mathbb{F}_{q^m}$.

For both of the above metrics d_* (where $* \in \{H, R\}$) and the corresponding codes C we define the *minimum distance* of C as

$$d_*(C) := \min\{d_*(u, v) \mid u, v \in C, u \neq v\}.$$

For both distances we consider additive error models, i.e., when sending a codeword $c \in C$ over the considered communication channel we receive

$$r = c + e,$$

for some error vector (or matrix) e that lives in the same space as c. We define the *weight* of a codeword to be its distance to the all-zero codeword. Then the distance between r and c is equal to the weight of e. If e is an error vector of weight at most $(d_*(C) - 1)/2$, then there is a unique closest codeword to r – namely c – with respect to the chosen metric. The process of finding the closest codeword c to a given received word r is called *(minimum distance) decoding*. It follows that we can correct (or decode) any error of weight at most $(d_*(C) - 1)/2$; hence we say that C has *error-correction capability* $\lfloor (d_*(C) - 1)/2 \rfloor$.

The various metrics for error correction arise from different applications with different noise behaviors. In general, given an application with a prescribed abstract error model, one tries to find a metric such that the most likely sent codeword is the closest codeword to a received word with respect to the metric. For example, in the classical telecommunication channel, where information is transmitted via periodic waveforms, the Hamming metric is used for *orthogonal modulation*, whereas the Lee metric is used for *phase modulation* [23]. In other applications we consider the complete loss of a symbol instead of added noise. In these *erasure channels* the Hamming metric is again the usual choice for recovering the codeword and the corresponding message.

One of the main applications for the rank metric is *linear network coding*. In this setting we consider broadcast communication over a network with one sender and several receivers who all want to get the same information. One can show that the capacity of such channels can be achieved by splitting the information into several packets which are simultaneously injected into the network via separate edges and letting the inner nodes of the network linearly combine their incoming information before forwarding it. The codewords can now be represented by stacking the packets that are sent simultaneously as rows in a matrix. However, the linear combinations inside the network may lead to a propagation of the errors during transmission over large parts of the network, which makes the Hamming metric unsuitable for this setting. It turns out that, if the operations of the network are known, the number of actual errors is best measured by the rank metric [3].

Many code constructions are known for the Hamming and the rank metric, of which we will use

- Reed-Solomon codes
- Goppa codes
- Gabidulin codes

in our lecture notes. We will define those in the following.

Definition 2.10. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_q$ be pairwise distinct. Moreover, let $v_1, \ldots, v_n \in \mathbb{F}_q^*$. Denote $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $v = (v_1, \ldots, v_n)$. Then

$$\mathrm{RS}_{n,k,q}(\alpha, v) := \{(v_1 f(\alpha_1), \ldots, v_n f(\alpha_n)) \mid f(x) \in \mathbb{F}_q[x], \deg f < k\}$$

is called a *generalized Reed-Solomon code* of length n and dimension k.

Theorem 2.11. *A generalized Reed-Solomon code of length n and dimension k has minimum Hamming distance $n - k + 1$ and is therefore an MDS (maximum distance separable) code.*[8]

Definition 2.12. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_{q^m}^*$ be pairwise distinct. Moreover, let $G(x) \in \mathbb{F}_{q^m}[x]$ be such that $G(\alpha_i) \neq 0$ and $v_i = \prod_{j \neq i}(\alpha_j - \alpha_i) G(\alpha_i)^{-1}$ for $i = 1, \ldots, n$. Denote $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $v = (v_1, \ldots, v_n)$. Then

$$RS_{n,k,q^m}(\alpha, v) \cap \mathbb{F}_q^n$$

is called a *q-ary Goppa code* of length n.

Note that with the definition above Goppa codes are subfield subcodes of Reed-Solomon codes. Originally (non-binary) Goppa codes were defined differently, however, the definitions are equivalent.

There is no closed formula for the dimension and the minimum Hamming distance of Goppa codes, however there are lower bounds on these parameters. For these lecture notes we do not need these lower bounds and refer the interested reader to [9].

Definition 2.13. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_{q^m}$ be linearly independent over $\mathbb{F}_q$ and write $\alpha = (\alpha_1, \ldots, \alpha_n)$. Denote by $\mathcal{L}_q[x]$ the set of $\mathbb{F}_q$-linearized polynomials[9] over $\mathbb{F}_{q^m}$. Then

$$\mathrm{Gab}_{n,k}(\alpha) := \{(f(\alpha_1), \ldots, f(\alpha_n)) \mid f(x) \in \mathcal{L}_q[x], \deg f < k\}$$

is called a *Gabidulin code* of length n and dimension k.

Theorem 2.14. *A Gabidulin code of length n and dimension k has minimum rank distance $n - k + 1$ and is therefore an MRD (maximum rank distance) code.*[10]

[8] MDS codes are optimal in the sense that they reach the upper bound $n-k+1$ on the minimum Hamming distance.

[9] Polynomials of the form $f(x) = \sum_i u_i x^{q^i}$ with $u_i \in \mathbb{F}_{q^m}$.

[10] MRD codes are optimal in the sense that they reach the upper bound $n-k+1$ on the minimum rank distance.

Lastly we need to know what the isometries (i.e., the distance preserving automorphisms) for the Hamming metric and the rank metric are. We denote by $\mathrm{GL}_n(q)$ the general linear group of invertible matrices in $\mathbb{F}_q^{n\times n}$ and by S_n the group of $n \times n$ permutation matrices.

Theorem 2.15. *1. The linear isometries for the metric space $(\mathbb{F}_q^n, d_H)$ are represented by the monomial matrices, i.e., matrices of the form*

$$\begin{pmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & v_n \end{pmatrix} P$$

for some $P \in S_n$ and non-zero $v_1, \dots, v_n \in \mathbb{F}_q$.

2. The linear isometries for the metric space $(\mathbb{F}_{q^m}^n, d_R)$ are represented by the elements of $\mathrm{GL}_n(q)$.

These linear isometries can be extended to semi-linear isometries by combining them (coordinate-wise) with the Galois group of $\mathbb{F}_q$, respectively $\mathbb{F}_{q^m}$.

Chapter 3

Code-Based Cryptosystems

3.1 General setup for public key encryption

A completely different application of error-correcting codes in any ambient metric space can be found in *code-based cryptography*, one of the main streams of *post-quantum cryptography*. Here the main focus is to secure data against eavesdroppers, in contrast to the channel coding setup above, where the reliability of data was the main objective.

The main idea of code-based cryptography is to use the *syndrome decoding problem (SDP)* as the underlying hard mathematical problem. The computational version of this problem can be stated as follows:

Syndrome Decoding Problem: Given some $t \in \mathbb{N}$, a parity check matrix $H \in \mathbb{F}_q^{r\times n}$ and a syndrome $s \in \mathbb{F}_q^r$, find a vector $e \in \mathbb{F}_q^n$ of weight at most t that fulfills $eH^\top = s$.

Note that the weight in our formulation can be any general weight; historically the SDP was first studied with respect to the Hamming metric, and thereafter for the rank metric and other metrics/weights. Its decisional version (i.e., the problem of deciding if such a vector e exists, without explicitly finding it) is known to be NP complete for any additive weight:

Theorem 3.1. *Let* $\mathrm{wt} : \mathbb{F}_q \to \mathbb{R}_{\geq 0}$ *be a function that satisfies the following properties:*

1. $\mathrm{wt}(0) = 0$,
2. $\mathrm{wt}(1) = 1$ *and* $\mathrm{wt}(x) \geq 1$ *for all* $x \neq 0$.

If wt *is extended additively to* $\mathbb{F}_q^n$ *(i.e., by adding the weights of the coordinates) then the SDP with respect to* wt *is NP-complete.*

Remark 3.2. Most known coding weights fulfill the conditions above, however, the rank metric does not, since it is not additive on the coordinates. It is currently

not known if the rank metric version of the SDP is NP-complete, however, there are indicators that it should be the case. (In particular, there is a probabilistic reduction of an NP-complete problem to the rank metric SDP.)

Remark 3.3. From a theoretical point of view it is one of the big selling points of code-based cryptography that the SDP is proven to be NP-complete. E.g., the integer factorization problem, which has been used in public-key cryptography in abundance over the last decades, is not NP-complete (under the assumption that N is not equal to NP).

There are two main general cryptosystems which are based on error-correcting codes – the *McEliece* [17] and the *Niederreiter system* [20]. Both of these have many variants, depending on which type of code one wants to use. Since both are equivalent from a security point of view (see [12]) we focus on the McEliece system in this proposal. For implementation purposes however, and for the construction of digital signatures, the Niederreiter system is of great interest, as well.

Originally, the McEliece cryptosystem was based on binary Goppa codes. The underlying idea in its general form, using an arbitrary linear block code, is as follows.

Algorithm 2 Generalized McEliece cryptosystem

- The receiver chooses a code C with generator matrix G and an efficient decoding algorithm. Moreover, they need a disguising function ϕ that is a near-isometry.[1]
- **Private key:** G and the decoding algorithm
- **Public key:** $\phi(G)$ together with the error correction capability of the code generated by $\phi(G)$, say $\hat{t}$
- **Encryption:** Choose a random error vector e of weight at most $\hat{t}$ and encrypt the message m as
$$c = m\ \phi(G) + e.$$
- **Decryption:** Compute $\phi^{-1}(c)$ and decode this in the secret code C to recover m.

Similarly the general framework of the Niederreiter system is described in Algorithm 3.

In both cryptosystems an attacker is not able to recover the message m without knowing ϕ, respectively the secret code C. As a brute force attack they can try to

[1] A near-isometry is a relaxation of the concept of isometry, namely a function on the vectors that changes the weight by at most a given value. To be precise, the function ϕ needs to have more properties than just being a near-isometry. However, we will not go into the technical details at this point.

Algorithm 3 Generalized Niederreiter algorithm

- The receiver chooses a code C with parity check matrix H and an efficient decoding algorithm. Moreover, they need a disguising function ϕ that is a near-isometry.
- **Private key:** H and the decoding algorithm
- **Public key:** $\phi(H)$ together with the error correction capability of the code which is the kernel of $\phi(H)$, say $\hat{t}$
- **Encryption:** Represent the message as an error vector e of weight at most $\hat{t}$ and encrypt it as
$$c = e\ \phi(H)^\top.$$
- **Decryption:** Compute $\phi^{-1}(c)$ and decode this in the secret code C to recover e and hence the message.

decode in the public code $\phi(C)$, but this code has no structure and hence no efficient decoding algorithm, and decoding in such a "random" code is too difficult.

Remark 3.4. One can show that the McEliece and the Niederreiter cryptosystems are equivalent from a security point of view, i.e., that if one can break one of them (with given parameters and codes) then one can also break the other.

The advantage of Niederreiter is that the ciphertext is smaller than in McEliece, whereas the disadvantage is that mapping the message to an error vector is not straight-forward.

For the above cryptosystems to be efficient and secure, the following have to be fulfilled:

- We need a code C that has efficient encoding and decoding algorithms.
- We assume that the code family used is known, hence we require this family to have enough elements to prevent brute force attacks.
- The error correction capability needs to be large enough such that a brute force attack on the possible errors can be prevented.
- We need a disguising function such that the original code cannot be found from the public generator matrix.
- Any generic decoding[2] should be infeasible.

[2]Generic decoding means a general decoding procedure that works for any (random or unstructured) code of the prescribed parameters.

Naturally, one could choose extremely long codes to increase the security parameters. However, this affects the efficiency and the key size of the cryptosystem. Generally, code-based cryptography suffers from large key sizes. To be secure against generic decoding attacks, the generator matrix of the secret code, and hence also the generator matrix of the public key, needs to be quite large. E.g., to achieve 96 bits security[3], we need approximately 10^6 bits in the public key size of the classical McEliece system. For comparison, RSA needs a public key of less than 2048 bits to achieve the same security level. On the other hand, the encryption and decryption times are very fast compared to other cryptosystems, which makes code-based cryptosystems very promising for many applications. One of the main research goals is hence to find codes and disguising functions that allow smaller key sizes than currently known variants.

Exercise 3.5. If $C \subseteq \mathbb{F}_q^n$ is a linear code of dimension k, what are the sizes of the public key and the ciphertext in the McEliece cryptosystem (in $q$bits or bits)? What are the respective sizes for the Niederreiter system? Can you compress the public key sizes by sharing more public information?

3.2 Overview of some variants

Over the last decades many variants of the McEliece (or Niederreiter) cryptosystem have been proposed. Recall that the original paper introducing the McEliece cryptosystem proposed to use binary Goppa codes. This system still remains unbroken. Since then, many other block codes have been studied in this system, e.g., Reed–Solomon, Reed–Muller, algebraic geometry, low density parity check (LDPC), wild Goppa, and polar codes. However, most of these codes are too structured, and the private key can be found efficiently by the attacker.

In the last years, some new code classes without known structural attacks were proposed to be used in code-based cryptosystems, e.g., quasi-cyclic MDPC (medium density parity check) codes [18]. This class results in significantly smaller key sizes compared to the original McEliece cryptosystem.

Another way of achieving smaller key sizes is to use rank metric codes, which was first suggested by Gabidulin, Paramonov, and Tretjakov (GPT) [6], since known generic decoding algorithms in the rank metric are less efficient than in the Hamming metric [1, 8]. Until recently, all proposed variants in the rank metric used Gabidulin codes; the main difference of the variants is the respective disguising function. Many of these variants have again been broken by now, mostly due to structural attacks, see e.g., [21, 11, 10]. However, recently some new variants were proposed, using other classes of codes, such as low rank parity check (LRPC) codes [7] or twisted Gabidulin codes [22]. Furthermore, other disguising functions have been proposed, e.g. in [14].

[3]Meaning that an attack needs at least 2^{96} operations.

In the following sections we will study some of these variants and explain their strengths and weaknesses.

3.3 Variants in the Hamming metric

3.3.1 The original McEliece system (Goppa codes)

As already mentioned, the original proposal by McEliece employs binary Goppa codes as secret codes. The generator matrix is then disguised by multiplying with an invertible matrix on the left (i.e., choosing a different basis) and multiplying with a permutation matrix on the right (i.e., applying an isometry on the code).

We state the complete description in Algorithm 4.

Algorithm 4 Original McEliece cryptosystem

- The receiver chooses a binary Goppa code $C \subseteq \mathbb{F}_2^n$ of dimension $k = n - mt$ with generator matrix G and minimum distance $2t + 1$. Moreover, they choose a random $A \in \mathrm{GL}_k(2)$ and $P \in S_n$.
- **Private key:** G and (A, P)
- **Public key:** $G' = AGP$
- **Encryption:** Choose a random error vector e of weight at most t and encrypt the message m as
$$c = m\,G' + e.$$
- **Decryption:** Compute $c' = cP^{-1}$ and decode this in C to recover mA. Recover m by multiplying with A^{-1}.

Let us now have a look at why this works: Note that in the decryption process we implicitly use the fact that

$$cP^{-1} = (mAGP + e)P^{-1} = \underbrace{mAG}_{\text{another codeword}} + \underbrace{eP^{-1}}_{\text{same weight as } e},$$

i.e., eP^{-1} is another uniquely decodable error and we hence recover the codeword mAG (and thus the corresponding message vector mA) with any decoder for C. Since A is known to the recipient, they can recover the original message m from mA.

Example 3.6. Bob chooses the small binary Goppa code of length 8 and dimension 2 with generator matrix

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

as secret code. This code has minimum distance 5 and can therefore correct $t = 2$ errors. Moreover, he randomly picks

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

Then he publishes (or sends to Alice) the public key

$$G' = AGP = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

Alice wants to send the message $m = \begin{pmatrix} 1 & 0 \end{pmatrix}$ and encrypts it as

$$\begin{aligned} c = mG' + e &= \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}, \end{aligned}$$

where e is randomly chosen among the vectors of weight at most $t = 2$. She sends c to Bob, who will then decrypt it by first inverting the operation of P,

$$c' = cP^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

and then decoding it to the closest codeword in C, which is

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

The corresponding message vector is $\begin{pmatrix} 1 & 0 \end{pmatrix}$ which leads to the message via

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

Remark 3.7. The above example is of course not secure since the code's parameters are very small. In general, the code parameters have to be chosen such that a generic decoder for the public code would not be able to find the message by just decoding in this (random-looking) public code.

It is a big part of code-based cryptographic research to determine the computational complexity of the best generic decoder. In the Hamming metric most of the best known decoders are *information set decoders (ISD)* or *birthday (paradox) decoders*. For an overview of such decoders we refer to [26].

Exercise 3.8. Set up a Niederreiter system with the same parameters as in Example 3.6. What is Alice's ciphertext and how would Bob recover the message?

3.3.2 McEliece systems based on Reed-Solomon codes

One question that easily comes up when considering the McEliece cryptosystem is why it uses binary Goppa codes. A natural idea is to use MDS (maximum distance separable) codes, due to their optimality and hence most compact representation for a given error correction capability. Therefore, we could easily replace the code C in Algorithm 4 with a (generalized) Reed-Solomon code. However, it turns out that Reed-Solomon codes give rise to a *distinguisher* (a certain type of key recovery) *attack*.

For this notice that Reed-Solomon codes distinguish themselves from random linear codes by the following property.

Definition 3.9. For $u, v \in \mathbb{F}_q^n$ define the *Schur/star/coordinate-wise product* as

$$u \star v := \left(u_1 v_1, \ldots, u_n v_n\right).$$

Then define the *(star) square code* of a linear code $C \subseteq \mathbb{F}_q^n$ to be

$$C^{\star 2} := \langle \{c \star d \mid c, d \in C\} \rangle.$$

Theorem 3.10. *[4, Theorem 2.3] If $C \subseteq \mathbb{F}_q^n$ is a random linear code then by high probability*

$$\dim(C^{\star 2}) = \min\left\{ \binom{k+1}{2}, n \right\},$$

whereas if C is a (generalized) Reed-Solomon code, then

$$\dim(C^{\star 2}) = \min\{2k - 1, n\}.$$

The above theorem presents an easily computable distinguisher for (generalized) Reed-Solomon codes. In fact, the (star product) square code can also be used for the following code families:

- low-codimensional subcodes of GRS codes,
- Reed-Muller codes,
- Polar codes,
- special types of Goppa codes,
- high rate alternant codes,
- algebraic geometry codes.

Now that we have a distinguisher it is still not straight-forward how to use this in a key recovery attack.

In Wieschebrink's paper [28], the star product is used to identify for a certain subcode C' of a GRS code a possible pair (x, y) of code locators and column multipliers. This is achieved by computing $C'^{\star 2}$ which turns out to be $C^{\star 2}$. The Sidelnikov-Shestakov algorithm [24] is then used on $C'^{\star 2}$ to recover suitable code locators and column multipliers of the original generalized Reed-Solomon code.

The idea of a distinguisher attack becomes more apparent when looking at Wieschebrink's cryptosystem [27] and the corresponding attack. Wieschebrink's system differs from the classical McEliece system with Reed-Solomon codes by using a different disguising function. We describe it in Algorithm 5.

Algorithm 5 Wieschebrink cryptosystem

- The receiver chooses a Reed-Solomon code $C \subseteq \mathbb{F}_q^n$ of dimension k with generator matrix G. They also choose $C_1, \ldots, C_r \in \mathbb{F}_q^k$, $S \in \mathrm{GL}_k(q)$ and $P \in S_{n+r}$ uniformly at random. Let $\bar{G}$ be the matrix obtained by concatenating G and the columns $C_1, \ldots, C_r$.
- **Private key:** G and (S, P)
- **Public key:** $G' = S^{-1}\bar{G}P^{-1}$
- **Encryption:** Choose a random error vector e of weight at most t and encrypt the message m as
$$c = m\, G' + e.$$
- **Decryption:** Compute $c' = cP^{-1}$, erase the last r coordinates and decode this in C to recover mS^{-1}. Recover m by multiplying with S.

To attack this system we use the fact that by puncturing the public generator matrix in random positions and computing the dimension of the star square code, we can identify the inserted columns $C_1, \ldots, C_r$. The attack goes as follows:

- Choose a random $1 \leq i \leq n+r$ and shorten G' in the ith position (i.e., erase the ith column in G').
- Compute the dimension of the square code of this new generator matrix.
- If the dimension of the square is

$$\begin{cases} 2k+r-2 & \text{then the erased column was probably from } \{C_1, \ldots, C_r\} \\ 2k+r-1 & \text{then the erased column was probably from the GRS code.} \end{cases}$$

- Identify all random columns like above to recover a generator matrix for the original GRS code.
- Apply Sidelnikov-Shestakov to recover the code locators and column multipliers to be able to decode.

Exercise 3.11. Prove Theorem 3.10.

Exercise 3.12. Prove that shortening a GRS code results in another GRS code. What are the dimension and the length of the shortened code?

Exercise 3.13. Prove that the square code of the public code generated by G' has dimension $2k+r$ by high probability.

3.3.3 McEliece system with LDPC/MDPC codes

A really different approach for disguising the private key can be done with low density parity check (LDPC) codes, as first proposed in [19]. In this case the decoding algorithm depends on a sparse parity check matrix of the code. We therefore do not need to hide the code used but only the sparse parity check matrix. We describe the algorithm in Algorithm 6.

Recall that the security of this system does not lie in hiding the code itself, but rather in the representation of the code that gives rise to an efficient decoder. Therefore, this system is not despicable to a distinguisher attack. However, there are key recovery attacks for this system, namely by finding low weight codewords in the dual code, which can then be used to design a sparse parity check matrix which in turn leads to an efficient decoder.

In return, when this attack based on finding low weight codewords was analyzed it was found that there is a regime of sparsity where the attack is not feasible any more, but the cryptosystem still works. The corresponding codes are now called *MDPC (medium/moderate density parity check) codes*[4] and the NIST standardization project finalist BIKE is partly based on this idea.

[4]Usually, a binary code is called MDPC if there exists a parity check matrix whose rows have weight $\mathcal{O}(\sqrt{n\log(n)})$.

Algorithm 6 McEliece with LDPC codes

- Choose an efficiently decodable LDPC code with sparse parity check matrix $H \subseteq \mathbb{F}_q^{r \times n}$.
- **Private key:** H
- **Public key:** G any generator matrix for the code defined by H
- **Encryption:** Choose a random error vector e of weight at most t and encrypt the message m as
$$c = m\,G + e.$$
- **Decryption:** Decode c in C (with the help of H) to recover m.

3.4 Variants in the rank metric

3.4.1 The original GPT system

The Gabidulin-Paramonov-Tretjakov (GPT) cryptosystem was introduced in [6] and is another variant of the McEliece cryptosystem, using codes and near-isometries for the rank metric instead of for the Hamming metric.

To understand the near-isometry that is used as the disguising function, we need to define one more concept:

Definition 3.14. 1. Let $X \in \mathbb{F}_{q^m}^{k \times n}$. We define the *column rank* of X to be the $\mathbb{F}_q$-dimension of the $\mathbb{F}_q$-space spanned by the columns of X.

2. Let $X \in \mathbb{F}_{q^m}^{k \times n}$ be a matrix of rank k and column rank t and $V \in \mathbb{F}_{q^m}^{k \times t}, U \in \mathbb{F}_q^{t \times n}$ such that $X = VU$. We call $\langle U \rangle$ the *Grassmann support* of X which will be denoted by $\langle U \rangle = \mathrm{supp}_{\mathrm{Gr}}(X)$.

Example 3.15. Consider $\mathbb{F}_4 = \mathbb{F}_2[\alpha]$ and the matrix

$$M = \begin{pmatrix} 1 & 0 & \alpha & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

in $\mathbb{F}_4^{2 \times 4}$. Its rank over $\mathbb{F}_4$ is 2, but its column rank is 3, since the first and third column are not $\mathbb{F}_2$-multiples of each other.

We can decompose M into

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

Then the Grassmann support of M is the row space of

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

The GPT cryptosystem is described in Algorithm 7.

Algorithm 7 GPT Cryptosystem

- The receiver chooses a Gabidulin code of minimum rank distance $2t+1$ with generator matrix $G \subseteq \mathbb{F}_{q^m}^{k \times n}$. Moreover, they randomly choose $S \in \mathrm{GL}_k(q^m)$ and $X \in \mathbb{F}_{q^m}^{k \times n}$ of column rank at most $s < t$.
- **Private key:** G and S
- **Public key:** $G' = SG + X$
- **Encryption:** Choose a random error vector e of rank weight at most $t - s$ and encrypt the message m as

$$c = m\, G' + e.$$

- **Decryption:** Decode c in the Gabidulin code to recover mS. Multiply with S^{-1} to recover m.

Similarly to Reed-Solomon codes, also Gabidulin codes suffer from distinguishing properties that make most of the cryptosystems using them vulnerable to key recovery attacks. Where we used the star product for distinguishing Reed-Solomon codes, Gabidulin codes can be distinguished by considering the intersection with itself under the coordinate-wise Frobenius map:[5]

Theorem 3.16. *[5] If $C \subseteq \mathbb{F}_{q^m}^n$ is a random linear code of dimension $0 < k < n$ then by high probability*

$$\dim(C \cap C^{(q)}) = \max\{2k - n, 0\},$$

whereas if C is a Gabidulin code, then

$$\dim(C \cap C^{(q)}) = k - 1.$$

Note that this behavior differs as soon as $1 < k < n - 1$, i.e., for any non-trivial Gabidulin code.

As before in the Hamming metric case, the question is now how to exploit this distinguishing property in an attack. We will explain such an attack in the following and last part of these lecture notes:

[5] This map – denoted by $x^{(q)}$ – raises every entry to its qth power when working over $\mathbb{F}_{q^m}$.

Distinguisher (key recovery) attack on the GPT cryptosystem based on rank 1 codewords

Consider the Gabidulin code $\mathrm{Gab}_{n,k}(\alpha)$ with dimension $1 < k < n$ and generator matrix SG, where $S \in \mathrm{GL}_k(\mathbb{F}_{q^m})$ and

$$G = \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \alpha_1^q & \alpha_2^q & \dots & \alpha_n^q \\ & & \vdots & \\ \alpha_1^{q^{k-1}} & \alpha_2^{q^{k-1}} & \dots & \alpha_n^{q^{k-1}} \end{pmatrix}. \tag{3.1}$$

Then $\mathrm{Gab}_{n,k}(\alpha)^{(q)} \cap \mathrm{Gab}_{n,k}(\alpha)$ is the Gabidulin code $\mathrm{Gab}_{n,k-1}(\alpha^{(q)})$. Iterating with this new Gabidulin code, we can eventually obtain a code of dimension 1, which is generated by $\alpha^{(q^{k-1})}$. If we take some non-zero element of this space, it has the form $\beta\alpha^{(q^{k-1})}$, for some $\beta \in \mathbb{F}_{q^m}$. Applying the Frobenius map coordinate-wise $m-k+1$ times, we obtain an element of the form $\beta^{q^{m-k+1}}\alpha$. Using this element, we can construct a generator matrix, BG, for $\mathrm{Gab}_{n,k}(\alpha)$ which will have the form

$$BG = \begin{pmatrix} \beta^{q^{m-k+1}} & & & \\ & \beta^{q^{m-k+2}} & & \\ & & \ddots & \\ & & & \beta \end{pmatrix} \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \alpha_1^q & \alpha_2^q & \dots & \alpha_n^q \\ & & \vdots & \\ \alpha_1^{q^{k-1}} & \alpha_2^{q^{k-1}} & \dots & \alpha_n^{q^{k-1}} \end{pmatrix}.$$

The change of basis from SG to BG is then given by BS^{-1}. For a message $m \in \mathbb{F}_{q^m}^k$, encoded as mSG, we can now decode with respect to $\mathrm{Gab}_{n,k}(\beta^{q^{m-k+1}}\alpha)$ to obtain mSB^{-1}. Then, applying BS^{-1}, we can recover m.

To set up our attack we must be able to find the elements of rank one in a linear rank metric code efficiently. To accomplish this, we only need to find the codewords that have all coordinates in $\mathbb{F}_q$ (all other rank one codewords are multiples of these). The following lemma shows how these codewords in $\mathbb{F}_q^n$ can be computed.

Lemma 3.17. *[10] Let $G \in \mathbb{F}_{q^m}^{k\times n}$ be in reduced row echelon form and denote by G_i the ith row of G. Then the solutions to*

$$\sum_{i=1}^{k} a_i(G_i^{(q)} - G_i) = 0, \tag{3.2}$$

for variables $a_i \in \mathbb{F}_q$, represent the codewords of $\langle G\rangle$ in $\mathbb{F}_q^n$.

When expanded over $\mathbb{F}_q$, Equation (3.2) gives rise to a linear system of equations with k variables, which can efficiently be solved with standard methods.

Now back to our attack to break the GPT cryptosystem, which extends Overbeck's attack [21] to cryptanalyze the system for all suggested parameters. Recall that the public key generator matrix is of the form

$$G' = SG + X \in \mathbb{F}_{q^m}^{k \times n},$$

where G is a generator matrix of a Gabidulin code $\mathrm{Gab}_{n,k}(\alpha)$, $X \in \mathbb{F}_{q^m}^{k \times n}$ is a matrix of column rank s, and $S \in \mathrm{GL}_k(q^m)$. For simplicity we will from now on assume that X cannot be decomposed into a Moore matrix[6] plus a matrix of less column rank. This assumption will simplify our explanation of the attack. The interested reader is referred to [11, 10] for a general explanation of the attack on any X, depending on the so-called *Moore decomposition* of X.

Note that, as an attacker, we do not have a priori knowledge of the parameter s. We can generally assume $s = t$, or else start with $s = 1$ and increase the value up to t until the attack succeeds.

We need one more preliminary lemma before getting to the main results used in our attack:

Lemma 3.18. *[10] Let $G \in \mathbb{F}_{q^m}^{k \times n}$ and X be as above. Then all elements of rank one in*

$$\sum_{i=0}^{s} \langle G + X \rangle^{(q^i)}$$

exactly span $\mathrm{supp}_{\mathrm{Gr}}(X)$.

Proof. Let $\mathcal{U}$ be the subspace spanned by all elements of rank one in

$$\sum_{i=0}^{s} \langle G + X \rangle^{(q^i)}.$$

Let $H \in \mathbb{F}_q^{(n-s) \times n}$ be a parity check matrix for $\mathrm{supp}_{\mathrm{Gr}}(X)$. We have

$$\begin{aligned} d_{\mathrm{R}}\left(\sum_{i=0}^{s} \langle G + X \rangle^{(q^i)} H^\top \right) &= d_{\mathrm{R}}\left(\sum_{i=0}^{s} \langle G \rangle^{(q^i)} H^\top \right) \\ &\geq 2t + 1 - 2s = 2(t - s) + 1 \geq 3. \end{aligned}$$

Since H is a matrix over $\mathbb{F}_q$, we get $\mathrm{wt}_{\mathrm{R}}(x) \geq \mathrm{wt}_{\mathrm{R}}(xH)$, and therefore we must have that all elements of rank one must be from a Frobenius power of X. It follows that $\mathcal{U} = \mathrm{supp}_{\mathrm{Gr}}(X)$. □

Theorem 3.19. *[10] Consider a GPT cryptosystem as defined above. Suppose an adversary can find a full rank matrix $U \in \mathbb{F}_q^{s \times n}$ satisfying*

$$\langle U \rangle = \langle X \rangle,$$

then an encrypted message can be recovered in polynomial time.

[6] I.e., a matrix of the form (3.1).

Proof. Let $H \in \mathbb{F}_q^{(n-s)\times n}$ be a parity check matrix for $\langle U \rangle$. Applying H to the public key generator matrix yields

$$G'H^\top = (SG + X)H^\top = SGH^\top.$$

Then, it follows that $\langle G \rangle H^T$ has minimum rank distance at least $n - k + 1 - s$. Moreover, $GH^\top$ is a Moore matrix.

From the minimum distance we know that there are $n - s$ independent columns in this matrix, which generate a Gabidulin code of minimum distance $n - s - k + 1$, $\mathrm{Gab}_{n-s,k}(\gamma)$, for some $\gamma \in \mathbb{F}_{q^m}^{n-s}$. From the results above, we can recover a decoding algorithm for $\mathrm{Gab}_{n-s,k}(\gamma)$ with respect to the submatrix formed by these $n - s$ columns. The error correction capability of $\mathrm{Gab}_{n-s,k}(\gamma)$ is

$$\left\lfloor \frac{n-s-k}{2} \right\rfloor = \left\lfloor t - \frac{s}{2} \right\rfloor \geq t - s \geq \mathrm{rank}(e) \geq \mathrm{rank}(eH^\top),$$

where the last inequality follows from the fact that H is a matrix over $\mathbb{F}_q$. For an encrypted message $m(SG + X) + e$, we have

$$(m(SG + X) + e)H^\top = mS(GH^\top) + eH^\top.$$

When we restrict this to the above chosen independent columns, we can uniquely decode in the respective code $\mathrm{Gab}_{n-s,k}(\gamma)$ and can therefore recover m. □

We can now use the previous result to attack and break the GPT cryptosystem.

Corollary 3.20. *[10] Consider a GPT cryptosystem as defined above with public key generator matrix $G' = SG + X \in \mathbb{F}_{q^m}^{k\times n}$. For any such cryptosystem, an encrypted message can be recovered in polynomial time.*[7]

Proof. We first note that

$$\left\lfloor \frac{n-k}{2} \right\rfloor = t > s.$$

By Corollary 3.18, all the elements of rank one in $\sum_{i=0}^{s} \langle G + S^{-1}X \rangle^{(q^i)}$ belong to $\mathrm{supp}_{\mathrm{Gr}}(S^{-1}X) = \mathrm{supp}_{\mathrm{Gr}}(X)$. With Lemma 3.17 we can find a basis matrix $U \in \mathbb{F}_q^{s\times n}$ for these elements of rank one in polynomial time. Then we can use Theorem 3.19 to recover the encrypted message. □

Exercise 3.21. Show that the decryption procedure in Algorithm 7 recovers the message m.

Exercise 3.22. The rank decomposition from Definition 3.14 is generally not unique. For a given rank decomposition, how can you create all other options? Why is the Grassmann support still well defined?

Exercise 3.23. Prove Lemma 3.17.

[7] The attack needs $O(k^2nm^2(s^2 + k))$ operations over $\mathbb{F}_q$, plus the operations needed for the Gabidulin code decoding algorithm. E.g., the decoding algorithm of [25] needs $O(m^3 \log m)$ operations over $\mathbb{F}_q$.

3.4.2 Other variants in the rank metric

There are many other variants in the rank metric, as already mentioned in the introduction to this chapter. They mostly differ in the choice of the disguising function (when using Gabidulin codes), or in using a similar idea as for LDPC codes, namely LRPC (low rank parity check) codes. We refer the interested reader to [26] and/or the references in Section 3.2 for further information on these systems.

3.5 Other metrics, other alphabets, other cryptosystems (outlook)

We gave a first overview of different variants of public key code-based encryption schemes, using codes and decoding algorithms in the Hamming or the rank metric. Historically, these were also the first two metrics studied in code-based cryptography. However, recently other metrics such as the sum-rank or the Lee metric have been studied for their use in McEliece type cryptosystems and promising first analyses were made. Furthermore, one needs not restrict to finite fields, but can also set up and analyze code-based cryptosystems over other alphabets, e.g., finite rings.

Remember that in the beginning we said that any public key encryption scheme can be transformed into a digital signature scheme. Theoretically, this is also true, but practically both the McEliece and the Niederreiter system suffer from the problem of turning the decoding procedure into a signing procedure. A few attempts have been made, but to no satisfactory result so far. It remains an interesting open question to design a practical code-based signature scheme (that is not of the form below).

Another avenue to design digital signatures is via a well-known transform – called the Fiat-Shamir transform – from zero knowledge identification schemes. There are code-based versions of these schemes and they can (practically) be used to create digital signatures. However, they suffer from large signature sizes, which is a problem that is inherent in the setup with zero knowledge identification schemes. Furthermore, it is questionable if these schemes are really code-based cryptography, since they do not make use of any decoding algorithms (which is at the heart of coding theory). On the other hand, they rely on the NP-hardness of the syndrome decoding problem and can therefore be called code-based.

References

[1] N. Aragon, P. Gaborit, A. Hauteville, and J.-P. Tillich. Improvement of generic attacks on the rank syndrome decoding problem. *Preprint, available at* `https://hal.archives-ouvertes.fr/hal-01618464`, 2017.

[2] E. R. Berlekamp, R. J. McEliece, and H. C. A. van Tilborg. On the inherent intractability of certain coding problems. *IEEE Trans. Inform. Theory*, IT-24(3):384–386, 1978.

[3] N. Cai and R. W. Yeung. Network error correction. II. Lower bounds. *Commun. Inf. Syst.*, 6(1):37–54, 2006.

[4] I. Cascudo, R. Cramer, D. Mirandola, and G. Zémor. Squares of random linear codes. *IEEE Trans. Inform. Theory*, 61(3):1159–1173, 2015.

[5] A. Couvreur, P. Gaborit, V. Gauthier-Umaña, A. Otmani, and J.-P. Tillich. Distinguisher-based attacks on public-key cryptosystems using Reed–Solomon codes. *Des. Codes Cryptogr.*, 73:641–666, 2014.

[6] E. M. Gabidulin, A. V. Paramonov, and O. V. Tretjakov. Ideals over a noncommutative ring and their application in cryptology. In *Advances in cryptology—EUROCRYPT '91 (Brighton, 1991)*, volume 547 of *Lecture Notes in Comput. Sci.*, pages 482–489. Springer, Berlin, 1991.

[7] P. Gaborit, G. Murat, O. Ruatta, and G. Zemor. Low Rank Parity Check codes and their application to cryptography. In L. Budaghyan, T. Helleseth, and M. G. Parker, editors, *The International Workshop on Coding and Cryptography (WCC 13)*, page 13 p., Bergen, Norway, Apr. 2013. ISBN 978-82-308-2269-2.

[8] P. Gaborit, O. Ruatta, and J. Schrek. On the complexity of the rank syndrome decoding problem. *IEEE Trans. Inform. Theory*, 62(2):1006–1019, 2016.

[9] V. D. Goppa. Algebraic-geometric codes. *Izv. Akad. Nauk SSSR Ser. Mat.*, 46(4):762–781, 896, 1982.

[10] A. Horlemann-Trautmann, K. Marshall, and J. Rosenthal. Considerations for rank-based cryptosystems. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT) 2016*, July 2016.

[11] A. Horlemann-Trautmann, K. Marshall, and J. Rosenthal. Extension of Overbeck's attack for Gabidulin based cryptosystems. *Des. Codes Cryptogr.*, 86:319–340, 2018.

[12] Y. X. Li, R. H. Deng, and X. M. Wang. On the equivalence of McEliece's and Niederreiter's public-key cryptosystems. *IEEE Transactions on Information Theory*, 40(1):271–273, 1994.

[13] R. Lidl and H. Niederreiter. *Introduction to Finite Fields and their Applications*. Cambridge University Press, Cambridge, London, 1994. Revised edition.

[14] P. Loidreau. A new rank metric codes based encryption scheme. In T. Lange and T. Takagi, editors, *Post-Quantum Cryptography*, pages 3–17, Cham, 2017. Springer International Publishing.

[15] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. I.* North-Holland Mathematical Library, Vol. 16. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977.

[16] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. II*. North-Holland Mathematical Library, Vol. 16. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977.

[17] R. J. McEliece. A public-key cryptosystem based on algebraic coding theory. Technical report, DSN Progress report # 42-44, Jet Propulsion Laboratory, Pasadena, Califonila, 1978.

[18] R. Misoczki, J.-P. Tillich, N. Sendrier, and P. S. Barreto. MDPC-McEliece: New McEliece variants from moderate density parity-check codes. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 2069–2073. IEEE, 2013.

[19] C. Monico, J. Rosenthal, and A. Shokrollahi. Using low density parity check codes in the McEliece cryptosystem. In *Proceedings of the 2000 IEEE International Symposium on Information Theory*, page 215, Sorrento, Italy, 2000.

[20] H. Niederreiter. Knapsack-type cryptosystems and algebraic coding theory. *Problems of Control and Information Theory 15*, 1(6):159–166, 1986.

[21] R. Overbeck. Structural attacks for public key cryptosystems based on Gabidulin codes. *J. Cryptology*, 21(2):280–301, 2008.

[22] J. R. S. Puchinger and A. Wachter-Zeh. Twisted Gabidulin codes in the GPT cryptosystem. In *Proceedings of International Workshop on Algebraic and Combinatorial Coding Theory (ACCT)*, 2018.

[23] M. Schwartz. *Information Transmission, Modulation, and Noise: A Unified Approach to Communication Systems*. McGraw-Hill, 1980.

[24] V. M. Sidelnikov and S. O. Shestakov. On an encoding system constructed on the basis of generalized Reed-Solomon codes. *Diskret. Mat.*, 4(3):57–63, 1992.

[25] A. Wachter-Zeh, V. Afanassiev, and V. Sidorenko. Fast decoding of Gabidulin codes. *Des. Codes Cryptogr.*, 66(1-3):57–73, 2013.

[26] V. Weger, N. Gassner, and J. Rosenthal. A survey on code-based cryptography. *https://arxiv.org/abs/2201.07119*, 2022.

[27] C. Wieschebrink. Two NP-complete problems in coding theory with an application in code based cryptography. In *2006 IEEE International Symposium on Information Theory*, pages 1733–1737, 2006.

[28] C. Wieschebrink. Cryptanalysis of the Niederreiter public key scheme based on GRS subcodes. In *PQCrypto 2010, Lecture Notes in Computer Science*, volume 6061, pages 61–72, 2010.

Part III

Points of Algebraic Varieties in generic position

Valentina Pepe

Dipartimento di Scienze di Base e Applicate per l'Ingegneria,
Sapienza University of Rome,
Via Scarpa 10,
00161 Rome,
Italy

email: valentina.pepe@sbai.uniroma1.it

Contents

Preface **73**

1 Introduction **75**

2 Some Remarkable Algebraic Varieties **77**

2.1 Exercises 84

3 An application to Coding Theory **85**

4 An application to Extremal Graph Theory **89**

4.1 An Open Problem 92

Bibliography **95**

Preface

The object of these Notes is the use of some classical algebraic varieties over finite fields to construct interesting codes or graphs. The literature about algebraic varieties over finite fields is quite wide, but we have tried to keep our approach as elementary as possible in order to reach as many combinatorialists as possible, as the main goal here is the application of some of their properties to Coding Theory and Extremal Graph Theory.

Chapter 1

Introduction

An (n, r)-set of $\mathrm{PG}(N-1, q)$ is a set consisting of n points such that any $r+1$ of them are linearly independent but some $r+2$ points are linearly dependent. Finding the maximum value $M_r(N, q)$ for which an (n, r)-set of $\mathrm{PG}(N-1, q)$ exists is a special case of the so called *packing problem*, see [20]. This is one of the oldest problem in Finite Geometry and it dates back to the seminal paper of Bose, see, e.g., [9], who first showed a connection of this problem with Coding Theory.

An $(n, N-1)$-set of $\mathrm{PG}(N-1, q)$ is called an *arc*. There is a wide literature about this topic, for a complete survey see [3]. There are applications of arcs in many fields, such as quantum physics (see, e.g. [7]), cryptography ([16]), group theory ([10]), but the best known application of arcs to the theory Error Correcting Codes, which is given by the well-established fact that a N-dimensional linear maximum distance separable code is equivalent to an arc in $\mathrm{PG}(N-1, q)$.

The topic was taken up by Beniamino Segre ([30, 29, 31, 33]): his fundamental work on arcs includes the celebrated result that a planar arc of $\mathrm{PG}(2, q)$ of size $q+1$ is a conic, that is a curve of degree 2. In fact, the most significant methodological contribution in Segre's work was to associate an algebraic curve to a planar arc and all the strongest results on arcs are based on Segre's initial ideas to find connections between arcs and algebraic curves/varieties. One fundamental open problem is whether there exist parameters (k, n) for which every arc of size $q+1$ of $\mathrm{PG}(k-1, q)$ is a normal rational curve. A partial answer to that was given in [2].

The aim of this notes is to show that a special class of algebraic varieties of $\mathrm{PG}(N-1, q)$ are (n, r)-sets quite "dense" in the relevant projective space, and how that can be used to construct interesting codes or graphs. Our starting point is, in fact, the algebraic variety $\mathcal{V}_{rt}$ of $\mathrm{PG}(r^t-1, q)$ introduced by Segre in [32], that is the Grassmann embedding of a Desarguesian $(t-1)$–spread of $\mathrm{PG}(rt-1, q)$. We stress out that a normal rational curve of $\mathrm{PG}(k-1, q)$ is the Grassmann embedding of a Segre variety Σ_{2k} (see [18]), and a Desarguesian $(t-1)$–spread of

$\mathrm{PG}(2t-1,q)$ is the union of Σ_{2t}. The algebraic variety $\mathcal{V}_{rt}$ is an $(\frac{q^{rt}-1}{q^t-1},t)$-set of $\mathrm{PG}(r^t-1,q)$. We present a generalization of that, namely the $(d,\boldsymbol{\sigma})$-Veronese variety, and some applications to Coding Theory and Extremal Combinatorics. We have tried too keep the approach as elementary and self-contained as possible. A good knowledge of tensor or wedge product could be useful but not necessary.

Chapter 3 is devoted to the introduction of the algebraic varieties we will use and the independence of their points. In Chapter 4 we show some application to Coding Theory and, finally, in Chapter 5, an application to a classical problem in Extremal Combinatorics.

Chapter 2

Some Remarkable Algebraic Varieties

We will follow [21] for notations and basic definitions.

Let $\mathbb{F}_q$ be the finite field of order q, $\mathrm{PG}(n-1,q)$ be the projective geometry over $\mathbb{F}_q$ of dimension $n-1$ and $\overline{\mathbb{F}_q}$ be the algebraic closure of $\mathbb{F}_q$. We shall identify a point $P \in \mathrm{PG}(n,q)$ with its homogeneous coordinate vector in $\mathbb{F}_q^n$. Let $\mathrm{PGL}(n,q)$, $\mathrm{P\Gamma L}(n,q)$ be, respectively, the projective linear and semi-linear groups of $\mathrm{PG}(n-,q)$.

By $\mathbb{F}_q \hookrightarrow \mathbb{F}_{q^t}$, we get a natural embedding of $\mathbb{F}_q^n$ in $\mathbb{F}_{q^t}^n$ and hence of $\mathrm{PG}(n-1,q)$ in $\mathrm{PG}(n-1,q^t)$. We say that $\mathrm{PG}(n-1,q)$ is a *subgeometry* of $\mathrm{PG}(n-1,q^t)$ of order q.

Throughout this paper we shall extensively use the following result: if σ is a $\mathbb{F}_q$–linear collineation of $\mathrm{PG}(n-1,q^t)$ of order t, then the subset $Fix(\sigma)$ of all elements of $\mathrm{PG}(n-1,q^t)$ point–wise fixed by σ is a subgeometry isomorphic to $\mathrm{PG}(n-1,q)$. This is a straightforward consequence of the fact that there is just one conjugacy class of $\mathbb{F}_q$–linear collineations of order t in $\mathrm{P\Gamma L}(n,q^t)$, namely that of $\mu : X \to X^q$. In particular, all subgeometries $\mathrm{PG}(n-1,q)$ are projectively equivalent to the set of fixed points of the map $(x_0,x_1,\ldots,x_{n-1}) \mapsto (x_0^q,x_1^q,\ldots,x_{n-1}^q)$.

Lemma 2.1 ([26, Lemma 1]). *Let $\Sigma \simeq \mathrm{PG}(n-1,q)$ be a subgeometry of $\mathrm{PG}(n-1,q^t)$ and let σ be the $\mathbb{F}_q$–linear collineation of order t such that $\Sigma = Fix(\sigma)$. Then a subspace Π of $\mathrm{PG}(n-1,q^t)$ is fixed set–wise by σ if and only if $\Pi \cap \Sigma$ has the same projective dimension as Π.*

Let $\mathbb{F}_q[x_0,x_1,\ldots,x_{n-1}]$ be the polynomial ring over $\mathbb{F}_q$ in n indeterminates. Let $f_i \in \mathbb{F}_q[x_0,x_1,\ldots,x_{n-1}]$ be a homogeneous polynomial and $\mathcal{V} := V(f_1,f_2,\ldots,f_r) = \{(a_0,a_1,\ldots,a_{n-1}) \in \mathrm{PG}(n,q) | f_i(a_0,a_1,\ldots,a_{n-1}) = 0 \,\forall\, i = 1,2,\ldots,r\}$. The set $\mathcal{V}$ is called the algebraic variety defined by $f_1,f_2,\ldots,f_r$ and, since $f_i \in \mathbb{F}_q[x_0,x_1,\ldots,x_{n-1}] \,\forall\, i = 1,2,\ldots,r$, we say that $\mathcal{V}$ is defined over $\mathbb{F}_q$. We might also look for the solutions

of $f_i(x_0, x_1, \ldots, x_{n-1}) = 0$ over some extension of $\mathbb{F}_q$ or over $\overline{\mathbb{F}_q}$, in that case we say that P is $\mathbb{F}_{q^t}$ or $\overline{\mathbb{F}_q}$-rational and we will denote by $\mathcal{V}_{\mathbb{F}}$ the set of all $\mathbb{F}$-rational solutions of $f_i(x_0, x_1, \ldots, x_{n-1}) = 0$, $i = 1, 2, \ldots, r$, with $\mathbb{F} = \mathbb{F}_{q^n}$ or $\overline{\mathbb{F}_q}$.

Let $J := \langle f_1, f_2, \ldots, f_r \rangle$ be the ideal of $\mathbb{F}_q[x_0, x_1, \ldots, x_{n-1}]$ generated by $f_1, f_2, \ldots, f_r$, then $V(f_1, f_2, \ldots, f_r) = V(J)$.

Let $I(\mathcal{V}) = \{f \in \mathbb{F}_q[x_0, x_1, \ldots, x_{n-1}] | f(a_0, a_1, \ldots, a_{n-1}) = 0 \, \forall \, (a_0, a_1, \ldots, a_{n-1}) \in \mathcal{V}\}$. The set $I(\mathcal{V})$ is an ideal of $\mathbb{F}_q[x_0, x_1, \ldots, x_{n-1}]$. For an ideal I of $\mathbb{F}[x_0, x_1, \ldots, x_{n-1}]$, the set $\sqrt{I} := \{f \in \mathbb{F}[x_0, x_1, \ldots, x_{n-1}] | f^m \in I$ for some integer $m \geq 1\}$ is called the radical of I and it is an ideal containing I. One important feature of the radical ideals is their connection with the ideals of algebraic varieties. The connection, however, is not shared whenever the field has positive characteristic.

Theorem 2.2 (Strong Nullstellenstatz, see, e.g., [6]). *Let $\mathbb{F}$ be an algebraic closed field and let J be an ideal of $\mathbb{F}[x_0, x_1, \ldots, x_{n-1}]$. Then $I(V(J)) = \sqrt{J}$.*

For finite fields, we have the following result.

Theorem 2.3. *Let J be an ideal of $\mathbb{F}_q[x_0, x_1, \ldots, x_{n-1}]$. Then $I(V(J)) = J + \langle x_0^q - x_0, x_1^q - x_1, \ldots, x_{n-1}^q - x_{n-1} \rangle$.*

An algebraic variety $\mathcal{V}$ of $\mathrm{PG}(n-1, q)$ is said to be *degenerate* if the span of $\mathcal{V}$ is a proper subset of $\mathrm{PG}(n-1, q)$.

The *Zariski topology* of a variety $\mathcal{V}$ is the topology on $\mathcal{V}$ whose closed sets are the subvarieties of $\mathcal{V}$, i.e., the common zero loci of polynomials on $\mathcal{V}$. Thus, a basis for the open subsets is given by the sets $U_f = \{P \in \mathcal{V} | f(P) \neq 0\}$ for f a homogeneous polynomial.

A common way to represent an algebraic variety is as the image of an embedding:

$$\epsilon : \mathrm{PG}(n-1, \mathbb{F}) \to \mathrm{PG}(N-1, \mathbb{F})$$

for some $N > n$, that is called in these notes, the *parametric representation* of an algebraic variety. Then, it is usually necessary to prove that $\mathrm{PG}(n-1, \mathbb{F})^\epsilon$ is the common zero loci of polynomials over $\mathbb{F}$.

We will briefly review a few remarkable examples.

The Normal Rational Curve

The rational normal curve $\mathcal{C} \in \mathrm{PG}(d, \mathbb{F})$ is defined to be the image of the map

$$v_d : \mathrm{PG}(1, \mathbb{F}) \longrightarrow \mathrm{PG}(d, \mathbb{F})$$

given by

$$v_d(x_0, x_1) \mapsto (x_0^d, x_0^{d-1}x_1, x_0^{d-2}x_2^2, \ldots, x_1^d) = (z_0, z_1, z_2, \ldots, z_d)$$

The image $\mathcal{C} \in \mathrm{PG}(d, \mathbb{F})$ is readily seen to be the common zero locus of the polynomials $F_{i,j} = z_i z_j - z_{i-1} z_{j+1}$ for $1 < i < j < d-1$. Note that for $d > 3$ it may also be expressed as the common zeros of a subset of these: the polynomials $F_{i,i}$, $i = 1, 2, \ldots, d-1$ and $F_{1,d-1}$, for example. We can easily list the set of points of $\mathcal{C}$:

$$\mathcal{C} = \{(1, t, t^2, \ldots, t^d), t \in \mathbb{F}\} \cup \{(0, 0, 0, \ldots, 1)\}.$$

The Veronese Variety

The construction of the rational normal curve can be further generalized: for any n and d, we define the Veronese map of degree d

$$v_d : \mathrm{PG}(n-1, \mathbb{F}) \longrightarrow \mathrm{PG}(N-1, \mathbb{F})$$

with $N = \binom{n-1+d}{d}$

by

$$v_d : (x_0, x_1, \ldots, x_{n-1}) \mapsto (\ldots, X_I, \ldots),$$

where X_I ranges over *all* the possible monomials of degree d in $x_0, x_1, \ldots, x_{n-1}$. It is not hard to see that the image of the Veronese map is an algebraic variety, often called the Veronese variety. Let H be a hyperplane of $\mathrm{PG}(N-1, \mathbb{F})$, then $H \cap \mathrm{PG}(n-1, \mathbb{F})^{v_d} = \{(x_0, x_1, \ldots, x_{n-1})^{v_d} | \sum_{I \in \mathcal{I}_{nd}} a_I X_I = 0\}$, where $\mathcal{I}_{nd} = \{(\alpha_0, \alpha_1, \ldots, \alpha_{n-1}) \in \mathbb{N}_0^n | \alpha_0 + \alpha_1 + \cdots + \alpha_{n-1} = d\}$ and $X_I = x_0^{\alpha_0} x_1^{\alpha_1} \cdots x_{n-1}^{\alpha_{n-1}}$, i.e., $H \cap \mathrm{PG}(n-1, \mathbb{F})^{v_d}$ is the Veronese embedding of the points that are the zeros of a polynomial of degree d. Viceversa, for every homogenous $f \in \mathbb{F}[x_0, x_1, \ldots, x_{n-1}]$ of degree d, $V(f)^{v_d}$ consists of the set of points of a suitable hyperplane section of $\mathrm{PG}(n-1, \mathbb{F})^{v_d}$.

The Segre Variety

The Veronese variety is, in turn, a subvariety of the Segre variety.

Let $\mathbf{x}^{(i)} := (x_0^{(i)}, x_1^{(i)}, \ldots, x_{n-1}^{(i)}) \in \mathbb{F}^n$, where $\mathbb{F}$ is any field. Then the Segre variety

$$\Sigma_{nd} := \underbrace{\mathrm{PG}(n-1, \mathbb{F}) \otimes \mathrm{PG}(n-1, \mathbb{F}) \otimes \cdots \otimes \mathrm{PG}(n-1, \mathbb{F})}_{d \text{ times}} \subset \mathrm{PG}(n^d - 1, \mathbb{F})$$

is the image of the map

$$s_d : \underbrace{\mathrm{PG}(n-1, \mathbb{F}) \times \mathrm{PG}(n-1, \mathbb{F}) \times \cdots \times \mathrm{PG}(n-1, \mathbb{F})}_{d \text{ times}} \to \mathrm{PG}(n^d - 1, \mathbb{F})$$

such that

$$s_d : (\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(d-1)}) \mapsto \left(\prod_{i=0}^{d-1} x_{f(i)}^{(i)} \right)_{f \in \mathfrak{F}}$$

where $\mathfrak{F}$ is the set of the maps $\{f : \{0, \ldots, d-1\} \to \{0, \ldots, n-1\}\}$, that is $\prod_{i=0}^{d-1} x_{f(i)}^{(i)}$ is monomial of degree d, product of d variables, each in a set $\{x_0^{(i)}, x_1^{(i)}, \ldots, x_{n-1}^{(i)}\}$, $i = 0, 1, \ldots, d-1$. We stress out that, since $(\mathbf{x}^{(i)}) \in \mathrm{PG}(n-1, \mathbb{F})$, $(\mathbf{x}^{(i)}) \neq \mathbf{0}$ $\forall\, i = 0, 1, \ldots, d-1$.

Hence, the Veronese maps v_d is the restriction of s_d on the vectors $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(d-1)})$ with $\mathbf{x}^{(0)} = \mathbf{x}^{(1)} = \ldots = \mathbf{x}^{(d-1)}$.

For those familiar with tensor products, the variety Σ_{nd} can be seen as the collection of the pure tensors $v_1 \otimes v_2 \otimes \cdots \otimes v_d$, with $v_i \in \mathbb{F}^n$. So, the veronese variety is the subvariety consisting of the pure tensors such that $v_1 = v_2 = \cdots = v_d$.

It is well known that Σ_{nd} is an algebraic variety and the Veronese variety is the intersection of Σ_{nd} with a suitable linear space of rank $N = \binom{n-1+d}{d}$.

Here an easy, explanatory example. The Segre variety Σ_{22} is the image of the map:

$$s_2 : ((x_0, y_0), (x_1, y_1)) \in \mathrm{PG}(1, \mathbb{F}) \times \mathrm{PG}(1, \mathbb{F}) \mapsto (x_0x_1, x_0y_1, y_0x_1, y_0y_1) \in \mathrm{PG}(3, \mathbb{F})$$

hence Σ_{22} is the hyperbolic quadrics of $\mathrm{PG}(3, \mathbb{F})$ of equation $z_0z_3 = z_1z_2$. If we restrict s_2 to the vectors (x_0, y_0, x_1, y_1) where $(x_1, y_1) = (x_0, y_0)$, we get the set $\mathcal{C} = \{(x_0^2, x_0y_0, y_0x_0, y_0^2), x_0, y_0 \in \mathbb{F}\}$, that is a conic which is the intersection of Σ_{22} with the plane of equation $z_1 = z = 2$.

The Grassmannian

Let $G(k, n)$ be the family of $(k-1)$-dimensional subspaces of $\mathrm{PG}(n-1, \mathbb{F})$, then the *Plüker embedding*

$$g_{k,n} : G(k, n) \to \mathrm{PG}(M-1, \mathbb{F})$$

with $M = \binom{n}{k}$, is defined by

$$g_{k,n} : \langle v_1, v_2 \ldots, v_k \rangle \mapsto v_1 \wedge v_2 \wedge \cdots \wedge v_k \in \mathrm{PG}(M-1, \mathbb{F})$$

where $\wedge$ is the wedge product and $\{v_1, v_2, \ldots, v_k\}$ is linearly independent. An elementary way to represent this embedding is the following. Let A be the $k \times n$ matrix over $\mathbb{F}$ whose rows are the vectors $v_1, v_2, \ldots, v_k$, then the Grassmann coordinates vector of $\langle v_1, v_2 \ldots, v_k \rangle$ is the vector consisting of all the minors of order k of A. Then the Plüker embedding maps a k-dimensional vector space to its Grassmann coordinates vector. The map is well defined, as if we pick another

set of vectors $\{v'_1, v'_2, \ldots, v'_k\}$ such that $\langle v'_1, v'_2, \ldots, v'_k\rangle = \langle v_1, v_2, \ldots, v_k\rangle$, then the two corresponding coordinates vectors are proportional.

The Grassmann variety is intersection of quadrics too.

The $(d, \boldsymbol{\sigma})$-Veronese variety

Let $G = Gal(\mathbb{F}_{q^t}|\,\mathbb{F}_q)$ be the *Galois group* of the Galois extension $\mathbb{F}_{q^t}/\mathbb{F}_q$ and $\boldsymbol{\sigma} = (\sigma_0, \ldots, \sigma_{d-1}) \in G^d$, $d \geq 1$, and define the map

$$\nu_{d,\boldsymbol{\sigma}} : v \in \mathrm{PG}(n-1, q^t) \longrightarrow v^{\sigma_0} \otimes v^{\sigma_1} \otimes \cdots \otimes v^{\sigma_{d-1}} \in \mathrm{PG}(n^d - 1, q^t)). \qquad (2.1)$$

Up to the action of the group $\mathrm{P\Gamma L}(n, q^t)$, we may assume that $\sigma_0 = 1$.

We will call $\nu_{d,\boldsymbol{\sigma}}$ the $(d, \boldsymbol{\sigma})$-*Veronese embedding* and, as defined before, its image $\mathcal{V}_{d,\boldsymbol{\sigma}}$ the $(d, \boldsymbol{\sigma})$-Veronese variety of *dimension* $n-1$. Then $\mathcal{V}_{d,\boldsymbol{\sigma}}$ is a variety of $\mathrm{PG}(N-1, \mathbb{F})$, $N = n^d$, with as many points as $\mathrm{PG}(n-1, q^t)$.

Since any element $\sigma_i \in G$ is a map of the type $\sigma_i : x \mapsto x^{q^{h_i}}$ with $0 \leq h_i < t$ and $0 \leq i \leq d-1$, hereafter we will suppose that

$$\boldsymbol{\sigma} = (\underbrace{\sigma_0, \ldots, \sigma_0}_{d_0 \text{ times}}, \underbrace{\sigma_1, \ldots, \sigma_1}_{d_1 \text{ times}}, \ldots, \underbrace{\sigma_m, \ldots, \sigma_m}_{d_m \text{ times}})$$

where $0 = h_0 < h_1 < \ldots < h_m < t$ and we will consider the vector $d_{\boldsymbol{\sigma}} = (d_0, d_1, \ldots, d_m)$ where d_j is the occurrence of σ_j in $\boldsymbol{\sigma}$, $0 \leq j \leq m$. Clearly $d_0 + d_1 + \ldots + d_m = d$. If $\boldsymbol{\sigma} \in G^d$, the integer

$$|\boldsymbol{\sigma}| = \sum_{i=0}^{d-1} q^{h_i} = \sum_{i=0}^{m} d_i q^{h_i}. \qquad (2.2)$$

will be called $norm$ of $\boldsymbol{\sigma}$.

Since we consider the ring of polynomials $\mathbb{F}_{q^t}[x_0, x_1, \ldots, x_{n-1}]$ actually as the quotient $\mathbb{F}_{q^t}[x_0, x_1, \ldots, x_{n-1}]/(x_0^{q^t} - x_0, x_1^{q^t} - x_1, \ldots, x_{n-1}^{q^t} - x_{n-1})$, **from now on we assume** $|\boldsymbol{\sigma}| < q^t$, so that distinct polynomials will be distinct functions over $\mathbb{F}_{q^t}$. By injectivity of map in (2.1), it is clear that $(d, \boldsymbol{\sigma})$-Veronese variety $\mathcal{V}_{d,\boldsymbol{\sigma}}$ has as many points as $\mathrm{PG}(n-1, q^t)$.

Example 2.4. Let $\boldsymbol{\sigma} = \mathbf{1}$, the identity of the product group G^d, the $(d, \boldsymbol{\sigma})$-Veronese variety $\mathcal{V}_{d,\boldsymbol{\sigma}}$ is the classical Veronese variety of degree d and $\mathcal{V}_{d,\boldsymbol{\sigma}} \subset \mathrm{PG}(N-1, q^t)$ with $N = \binom{n+d-1}{d}$.

Example 2.5. Let σ be a generator of $Gal(\mathbb{F}_{q^t}|\mathbb{F}_q)$ and $\boldsymbol{\sigma} = (1, \sigma, \ldots, \sigma^{t-1})$, then we get the algebraic variety introduced in [32, 26, 28] and we will refer to it as the Segre-Lunardon-Pepe (SLP for short) -*variety* $\mathcal{V}_{t,\sigma}$. Hence,

$$\mathcal{V}_{t,\sigma} = \{v \otimes v^{\sigma} \otimes \cdots \otimes v^{\sigma^{t-1}}, v \in \mathbb{F}_{q^t}^n \setminus \{\mathbf{0}\}\} \in \mathrm{PG}(n^t - 1, q^t).$$

In this case, in fact, $\mathcal{V}_{t,\sigma}$ turns out to be a variety of a subgeometry $\mathrm{PG}(n^t-1,q) \subset \mathrm{PG}(n^t-1,q^t)$ in the following way.

Let $\{e_0, e_1, \ldots, e_{n-1}\}$ be a basis for $\mathbb{F}^n$, then it is well known that $\{e_{i_0} \otimes e_{i_1} \otimes \cdots \otimes e_{i_{t-1}}, i_j \in \{0,1,\ldots,n-1\}$ is a basis for $\mathbb{F}^{n^t}$ for any field $\mathbb{F}$.

Let ϕ_j, $j = 0,1,\ldots,t-1$, be collineations of $\mathrm{P\Gamma L}(n,q^t)$ with the same companion field automorphism and let f be a permutation of $\{0,1,\ldots,t-1\}$. Then the map:

$$e_{i_0} \otimes e_{i_1} \otimes \cdots \otimes e_{i_{t-1}} \mapsto e_{i_{f(0)}}^{\phi_0} \otimes e_{i_{f(1)}}^{\phi_1} \otimes \cdots \otimes e_{i_{f(t-1)}}^{\phi_{t-1}}$$

can be extended by linearity to a collineation of $\mathrm{PG}(n^t-1,q^t)$ and hence induce the map:

$$\hat{\phi} : v_0 \otimes v_1 \otimes \cdots \otimes v_{t-1} \mapsto v_{f(0)}^{\phi_0} \otimes v_{f(1)}^{\phi_1} \otimes \cdots v_{f(t-1)}^{\phi_{t-1}}$$

on Σ_{nt}.

Therefore, there exits an $\mathbb{F}_q$-linear collineation $\hat{\sigma}$ of $\mathrm{PG}(n^t-1,q)$ acting on Σ_{nt} in the following way:

$$v_0 \otimes v_1 \otimes \cdots \otimes v_{t-1} \mapsto v_{t-1}^{\sigma} \otimes v_0^{\sigma} \otimes \cdots \otimes v_{t-2}^{\sigma}.$$

The semi-linear collineation $\hat{\sigma}$ has order t, hence $Fix(\hat{\sigma}) \simeq \mathrm{PG}(n^t-1,q)$ and clearly $\mathcal{V}_{t,\sigma} \subset Fix(\hat{\sigma})$.

We have that, not only $\mathcal{V}_{t,\sigma}$ is a variety of $\mathrm{PG}(n^t-1,q)$, but also it is the Grassmann embedding of the elements of a Desarguesian $(t-1)$–spread of $\mathrm{PG}(nt-1,q)$ (see [32, 26]).

By (2.1), a point of $\mathrm{PG}(n-1,q^t)$ with homogeneous coordinates $(x_0,x_1,\ldots,x_{n-1})$ is mapped by $\nu_{d,\boldsymbol{\sigma}}$ into a point of coordinates

$$(\ldots, \prod_{j=0}^{m} X_{I_j}^{\sigma_j}, \ldots)$$

where X_{I_j} is a monomial of degree d_j in the variables $x_0,x_1,\ldots,x_{n-1}$. Hence, the $(d,\boldsymbol{\sigma})$-Veronese variety $\mathcal{V}_{d,\boldsymbol{\sigma}}$ is contained in a projective space of vector space dimension

$$N = N_0 N_1 \cdots N_m, \quad N_j = \binom{n+d_j-1}{d_j}, \quad j = 0,1,\ldots,m. \tag{2.3}$$

In the following, we just show under which hypothesis $\mathcal{V}_{d,\boldsymbol{\sigma}}$ is non-degenerate.

Theorem 2.6. *[12] Let $\boldsymbol{\sigma} \in G^d$ with $d_{\boldsymbol{\sigma}} = (d_0, d_1, \ldots, d_m)$, $|\boldsymbol{\sigma}| < q^t$. The $(d, \boldsymbol{\sigma})$-Veronese variety $\mathcal{V}_{d,\boldsymbol{\sigma}}$ is not contained in any hyperplane of $\mathrm{PG}(N-1, q^t)$ with $N = N_0 N_1 \cdots N_m$ and*

$$N_j = \binom{n + d_j - 1}{d_j}, \quad j = 0, 1, \ldots, m.$$

We will briefly recall the definition of *dimension* and *degree* of an algebraic variety as we will need them in the last Chapter, although they do not fit well for finite fields.

When we say that the *general* $(k-1)$-subspace of $\mathrm{PG}(n-1, \mathbb{F})$ has the property $\mathcal{P}$, we mean that the points of $\mathcal{G}_{nk}$ corresponding to the subspaces with the property $\mathcal{P}$ form a Zariski open dense subset of $\mathcal{G}_{nk}$.

Definition 2.7. The dimension of a projective variety $\mathcal{V} \in \mathrm{PG}(n-1, \mathbb{F})$ is the smallest integer k such that there exists a subspace of dimension $n-k-2$ of $\mathrm{PG}(n-1, \overline{\mathbb{F}})$ disjoint from $\mathcal{V}_{\overline{\mathbb{F}}}$.

Or, equivalently:

Definition 2.8. The dimension of a projective variety $\mathcal{V} \in \mathrm{PG}(n-1, \mathbb{F})$ is that integer k such that the general $(n-k-1)$-subspace in $\mathrm{PG}(n-1, \overline{\mathbb{F}})$ intersects $\mathcal{V}_{\overline{\mathbb{F}}}$ in a finite set of points.

Definition 2.9. Let $\mathcal{V}$ be a projective variety of $\mathrm{PG}(n-1, \mathbb{F})$ of dimension k, then the degree of $\mathcal{V}$ is the number of points of intersection of $\mathcal{V}_{\overline{\mathbb{F}}}$ with a general $(n-k)$-subspace of $\mathrm{PG}(n-1, \overline{\mathbb{F}})$.

Example 2.10. A conic is a curve, that is an algebraic variety of dimension 1, of degree 2, but there are "plenty" of lines of $\mathrm{PG}(2, q)$ disjoint from a conic over a finite field.

A topic of great interest and with many application is the estimation of the number of points of an algebraic variety over a finite field. We will not need that in our notes, as the varieties we are considering are image of an injective map, so we know exactly how many points they have. Nevertheless, it is worth mention (and we will in fact use that in the last Chapter) that a variety of dimension k over $\mathbb{F}_q$ as *roughly* q^k points (see, e.g. [25]).

We conclude this Chapter with an overview of the independence property of the points of the aforementioned varieties.

Let $\mathcal{C}$ be the normal rational curve of $\mathrm{PG}(d, q)$, i.e., up the action of $\mathrm{PGL}(d+1, q)$,

$$\mathcal{C} = \{(1, t, t,^2, \ldots, t^d), t \in \mathbb{F}_q\} \cup \{(0, 0, \ldots, 1)\}$$

and if $q \geq d$, $\mathcal{C}$ is non-degenerate. One can easily see that any $d+1$ points of $\mathcal{C}$ are linearly independent, for example, by considering the matrix A whose rows are the coordinates vectors of $d+1$ distinct points of $\mathcal{C} \setminus \{(0,0,\ldots,1)\}$: A turns out to be a Vandermonde matrix with distinct rows, hence $\det(A) \neq 0$. Since $\mathcal{C} \subset \mathrm{PG}(d,q)$, any $d+2$ of $\mathcal{C}$ are linearly dependent, hence $\mathcal{C}$ is an $(q+1,d)$-set of $\mathrm{PG}(d,q)$.

That can be generalized to Veronese varieties. Also here we assume $q \geq d$. Let $\{P_0, P_1, \ldots, P_d\}$ be $d+1$ distinct points of $\mathrm{PG}(n-1,q)^{v_d}$, hence $P_i = p_i^{v_d}$ for some $p_i \in \mathrm{PG}(n-1,q)$. Let $a_0^{(i)}x_0 + a_1^{(i)}x_1 + \cdots + a_{n-1}^{(i)}$ be a linear function vanishing in p_i and not in p_0, $j \neq i$, for $i = 1,2,\ldots,d$ (does that always exist?). Then $\prod_{i=1}^{d}(a_0^{(i)}x_0 + a_1^{(i)}x_1 + \cdots + a_{n-1}^{(i)})$ is a polynomial of degree d vanishing in $p_1, p_2, \ldots, p_d$ but not in p_0, hence it corresponds to a hyperplane section of $\mathrm{PG}(n-1,q)^{v_d}$ containing $\{P_1, P_2, \ldots, P_d\}$ but not P_0. Therefore, by the generality of P_0, we can say that $\{P_0, P_1, \ldots, P_d\}$ is linearly independent. Since for a line ℓ of $\mathrm{PG}(n-1,q)$, ℓ^{v_d} is a normal rational curve, we know that there exists $d+2$ points of $\mathrm{PG}(n-1,q)^{v_d}$ linearly dependent, hence $\mathrm{PG}(n-1,q)^{v_d}$ is an $\left(\frac{q^n-1}{q-1}, d\right)$-set of $\mathrm{PG}(\binom{n+d-1}{d}-1, q)$.

In a complete analogue way, one can prove the same result for $\mathcal{V}_{d,\boldsymbol{\sigma}}$.

Let $\{P_0, P_1, \ldots, P_d\}$ be $d+1$ distinct points of $\mathcal{V}_{d,\boldsymbol{\sigma}}$, hence $P_i = p_i^{\nu_{d,\boldsymbol{\sigma}}}$, $i = 0,1,\ldots,d$. Let $l^{(i)}(x_0, x_1, \ldots, x_{n-1}) := a_0^{(i)}x_0 + a_1^{(i)}x_1 + \cdots + a_{n-1}^{(i)}$ be a linear function vanishing in $p_i^{\sigma_i}$ and not in $p_0^{\sigma_i}$, $i = 1,2,\ldots,d$, then $\prod_{i=1}^{d} l_i(p_i^{\sigma_i}) = 0$ defines a hyperplane of $\mathcal{V}_{d,\boldsymbol{\sigma}}$ containing $\{P_1, P_2, \ldots, P_d\}$ but not P_0. Hence, any $d+1$ distinct points of $\mathcal{V}_{d,\boldsymbol{\sigma}}$ are linearly independent.

2.1 Exercises

Exercise 2.11. Let $S = \{v_1^{(i)} \otimes v_2^{(i)} \otimes \cdots \otimes v_d^{(i)}, i = 0,1,\ldots,d\}$ be a set of points of Σ_{nd} such that $v_i^{(j_1)} \neq v_i^{(j_2)}$ for every $j_1 \neq j_2$. Show that S is linearly independent. Deduce from that the linear independence of the points of $\mathcal{V}_{d,\boldsymbol{\sigma}}$.

Exercise 2.12. Let $\mathcal{V}$ be the image of the map $\varepsilon : \mathrm{PG}(n-1,\mathbb{F}) \to \mathrm{PG}(\binom{n}{d}-1,\mathbb{F})$ such that

$$\varepsilon : (x_0, x_1, \ldots, x_{n-1}) \mapsto \left(\prod_{i \in I} x_i\right)_{I \subset \{0,1,\ldots,n-1\} || I| = d}.$$

under which hypothesis we can say that any $d+1$ points of $\mathcal{V}$ are linearly independent?

Chapter 3

An application to Coding Theory

A $[\nu, \kappa]$-*linear code* $\mathcal{C}$ is a subspace of the vector space $\mathbb{F}_q^\nu$ of dimension κ. The *weight* of a codeword is the number of its entries that are nonzero and the *Hamming distance* between two codewords is the number of entries in which they differ. The distance δ of a linear code is the minimum distance between distinct codewords and it is equals to the minimum weight. A linear code of length ν, dimension κ, and minimum distance δ is called a $[\nu, \kappa, \delta]$-code. A matrix H of order $(\nu - \kappa) \times \nu$ such that

$$\mathbf{x}H^T = \mathbf{0} \quad \text{for all } \mathbf{x} \in \mathcal{C}$$

is called a *parity check matrix* for $\mathcal{C}$. The minimum weight, and hence the minimum distance, of $\mathcal{C}$ is at least w if and only if any $w-1$ columns of H are linearly independent [27, Theorem 10, p. 33]. Each linear $[\nu, \kappa, \delta]$-code $\mathcal{C}$ satisfies the following inequality

$$\delta \leq \nu - \kappa + 1,$$

called *Singleton bound*. If $\delta = \nu - \kappa + 1$, $\mathcal{C}$ is called *maximum distance separable* or *MDS*, while if $\delta = \nu - \kappa$ the code is called *almost MDS*. These can be related to some subsets of points in the projective space. More precisely, $\mathcal{C}$ is a $[\nu, \kappa, \delta]$-linear code if and only if the columns of its parity check matrix H can be seen as ν points in $\mathrm{PG}(\nu-\kappa-1, q)$ each $\delta-1$ of which are in general position, [11, Theorem 1]. Then, the existence of MDS or almost MDS linear codes is equivalent to the existence of arcs or (n, r)-sets in projective spaces.

If H is the matrix whose columns are the coordinates vectors of the points of the variety $\mathcal{V}_{d,\boldsymbol{\sigma}}$, we get a code $\mathcal{C}_{d,\boldsymbol{\sigma}}$ and we may study the minimum distance of it and characterize the codewords of minimum weight.

Definition 3.1. Let $\mathcal{V}_{d,\boldsymbol{\sigma}}$ be a $(d, \boldsymbol{\sigma})$-Veronese variety and denote by $\mathcal{C}_{d,\boldsymbol{\sigma}}$ the code whose parity check matrix H of order $N \times (\frac{q^{nt}-1}{q^t-1})$ has columns that are the coordinate vectors of the points of the variety $\mathcal{V}_{d,\boldsymbol{\sigma}}$.

Let N be defined as in 2.3 and let $|\boldsymbol{\sigma}| < q^t$, then, by Theorem 2.6, $\mathcal{C}_{d,\boldsymbol{\sigma}}$ is $\left[\frac{q^{nt}-1}{q^t-1}, \frac{q^{nt}-1}{q^t-1} - N\right]$-code. Also, we have showed that any $d+1$ points of $\mathcal{V}_{d,\boldsymbol{\sigma}}$, hence $\mathcal{C}_{d,\boldsymbol{\sigma}}$ is an $\left[\frac{q^{nt}-1}{q^t-1}, \frac{q^{nt}-1}{q^t-1} - N, d+2\right]$-code. As a matter of fact, to show that the minimum distance is exactly $d+2$ we need to show that there do exist $d+2$ points of $\mathcal{V}_{d,\boldsymbol{\sigma}}$ which are linearly dependent. In [12] we prove that, actually characterizing the sets of $d+2$ points linearly dependent.

Theorem 3.2. *[12] A set of $d+2$ linearly dependent points of $\mathcal{V}_{d,\boldsymbol{\sigma}}$ is the $\boldsymbol{\sigma}$-Veronese embedding of points on a subline $\cong \mathrm{PG}(1,q')$, where $\mathbb{F}_{q'}$ is the largest subfield of $\mathbb{F}_{q^t}$ fixed by σ_i in $\boldsymbol{\sigma}$.*

So we are able to characterize the minimum weight codewords of $\mathcal{C}_{d,\boldsymbol{\sigma}}$.

Definition 3.3. The support of a codeword $\mathbf{w} \in \mathcal{C}_{d,\boldsymbol{\sigma}}$ is the set of the points of the variety $\mathcal{V}_{d,\boldsymbol{\sigma}}$ corresponding to the non-zero positions of $\mathbf{w}$.

Theorem 3.4. *A codeword $\mathbf{w} \in \mathcal{C}_{d,\boldsymbol{\sigma}}$ has minimum weight if and only if its support consists of $d+2$ points contained in the image of a subline $\mathrm{PG}(1,q')$, $d < q'$, where $\mathbb{F}_{q'}$ is the largest subfield of $\mathbb{F}_{q^t}$ fixed by σ_i for all σ_i in $\boldsymbol{\sigma}$.*

As we have seen in Example 2.5, the SLP-variety turns out to be a variety of a subgeometry of order q, even though the array $\boldsymbol{\sigma}$ is defined on a finite field of order q^t, hence among all the possible choice of $\boldsymbol{\sigma}$ and n, for q 'big enough' $\mathcal{V}_{t,\sigma}$ is the variety with the most 'dense' set of points of a projective space with the property that any $d+1$ points are independent and therefore $\mathcal{C}_{d,\boldsymbol{\sigma}}$ is a more efficient code.

On the other hand, it is worth investigating these codes for small q.

Let N be as in (2.3) with $\sum_{i=0}^{m} d_i = d$. If $n = 2$, then

$$N = \prod_{i=0}^{m} (d_i + 1)$$

and the minimum is reached for $m = 1, d_0 = d-1, d_1 = 1$, so $N = 2d$.
If σ is such that $Fix(\sigma) \cap \mathbb{F}_{q^t} = \mathbb{F}_p$, where p is the characteristic of the field, since we should have $d \geq p$, the smallest possible $d = p$ and in this case

$$\boldsymbol{\sigma} = (\underbrace{1, 1, \ldots, 1}_{p-1 \text{ times}}, \sigma) \tag{3.1}$$

getting that $\mathcal{V}_{d,\boldsymbol{\sigma}}$ is a set of q^t+1 points in $\mathrm{PG}(2p-1, q^t)$ such that any $p+2$ of them are in general position. So the code $\mathcal{C}_{d,\boldsymbol{\sigma}}$ is a $[q^t+1, q^t-2p+1]$-linear code with minimum distance at least $p+3$ and the Singleton bound $2p+1$. Now, if $\sigma : x \mapsto x^p$, then $\mathcal{V}_{p,\boldsymbol{\sigma}}$ is the normal rational curve of $\mathrm{PG}(2p-1, q^t)$; hence $\mathcal{C}_{p,\boldsymbol{\sigma}}$ is an MDS code.
Furthermore for $p \in \{2, 3\}$, the following cases can also occur

- for $p = 2$, $\sigma : x \mapsto x^{2^h}$, $1 < h < et$, $\mathcal{V}_{2,\boldsymbol{\sigma}}$ is either the Segre arc or the normal rational curve (for $h = et - 1$), hence $\mathcal{C}_{2,\boldsymbol{\sigma}}$ is an MDS code.
- for $p = 3$, $\sigma : x \mapsto x^{3^h}$, $1 < h < et$, $\mathcal{V}_{3,\boldsymbol{\sigma}}$ is a $(3^{et}+1)$-track of $\mathrm{PG}(5, 3^{et})$; hence $\mathcal{C}_{3,\boldsymbol{\sigma}}$ is a so called *almost MDS* code, [11], see next Theorem 3.6.

Clearly, as p gets larger, the minimum distance gets smaller than the Singleton bound. Before showing the announced result, we recall the following theorem due to Thas [34] and of which Kaneta and Maruta gives an elementary proof,

Theorem 3.5. *[22, Theorem 1] In $\mathrm{PG}(r, q)$, $r \geq 2$ and q odd, every k-arc with*

$$q - \sqrt{q}/4 + r - 1/4 \leq k \leq q + 1$$

is contained in one and only one normal rational curve of the space. In particular, if $q > (4r-5)^2$, then every $(q+1)$-arc is a normal rational curve.

Theorem 3.6. *Let $q = 3^e$ and $\sigma : x \in \mathbb{F}_{q^t} \mapsto x^{3^h} \in \mathbb{F}_{q^t}$, $1 < h < et$, $\gcd(h, et) = 1$ with $et > 4$. Then $\mathcal{V}_{3,\boldsymbol{\sigma}}$ with $\boldsymbol{\sigma} = (1, 1, \sigma)$ is a $(3^{et}+1)$-track of $\mathrm{PG}(5, 3^{et})$ and $\mathcal{C}_{3,\boldsymbol{\sigma}}$ is an almost MDS.*

Proof. By the previous considerations, since the $[q^t+1, q^t-5]$-code $\mathcal{C}_{d,\boldsymbol{\sigma}}$ has distance at least 6, the result follows showing the existence of 6 columns of H linearly dependent or equivalently that there exists 6 points linearly dependent of the set

$$\mathcal{V}_{3,\boldsymbol{\sigma}} = \{(1, z, z^2, z^{3^h}, z^{3^h+1}, z^{3^h+2}) \,:\, z \in \mathbb{F}_{q^t}\} \cup \{(0,0,0,0,0,1)\}.$$

Suppose that any 6 points of $\mathcal{V}_{3,\boldsymbol{\sigma}}$ with $\boldsymbol{\sigma} = (1, 1, \sigma)$ are linearly independent, hence $\mathcal{V}_{3,\boldsymbol{\sigma}}$ is an arc of $\mathrm{PG}(5, q^t)$. By Theorem 3.5, $\mathcal{V}_{3,\boldsymbol{\sigma}}$ must be projectively equivalent to rational normal curve

$$\{(1, y, y^2, y^3, y^4, y^5) \,:\, y \in \mathbb{F}_{q^t}\} \cup \{(0,0,0,0,0,1)\}.$$

Since the normal rational curve has a 3-transitive automorphisms group, we can always assume that there is a collineation of $\mathrm{PG}(5, q^t)$ fixing $(0,0,0,0,0,1)$ and $(1,0,0,0,0,0)$. Moreover, w.l.o.g. we can assume that this collineation has the identity as companion automorphism.

Hence there must be $f_i(y) \in \mathbb{F}_{q^t}[y]$ of degree at most 5 and linearly independent such that

$$(f_0(y), f_1(y), f_2(y), f_3(y), f_4(y), f_5(y)) = (1, z, z^2, z^{3^h}, z^{3^h+1}, z^{3^h+2})$$

with $f_i(y)$ vanishing in 0 for $i \in \{1,2,3,4,5\}$ and $f_0(0) = 1$ up to a nonzero scalar. So, $f_0(y) = 1$ for all $y \in \mathbb{F}_{q^t}$ and since $\deg f_0(y) \leq 5 < q^t$, then $f_0(y) = 1$. Note that $\deg f_i(y) \neq 0$ for $i = 1, 2, 3, 4, 5$ and

$$f_2(y) = f_1(y)^2 \mod y^{q^t} - y,$$

but $2\deg f_1(y) \leq 10 < q^t$, and hence $f_2(y) = f_1(y)^2$ and $\deg f_1(y) \leq 2$. Similarly,

$$f_4(y) = f_1(y)^{3^h} \mod y^{q^t} - y,$$

but $3^h \deg f_1(y) \leq 3^h \cdot 2 < q^t$, so $f_4(y) = f_1(y)^{3^h}$ and $3^h \deg f_1(y) \leq 5$, obtaining $3^h \leq 5$, a contradiction. □

Actually, the result above holds for $q^t = 27, 81$ as well, this is verified by the software MAGMA, obtaining an infinite family of almost MDS codes or, equivalently, an infinite family of $(3^{et}+1)$-tracks of $\mathrm{PG}(5, 3^{et})$ with $et > 2$.

Chapter 4

An application to Extremal Graph Theory

One of the most famous problem in Extremal Graph Theory is the *Forbidden Subgraph Problem*.

For a graph H, what is the maximal number $ex(n, H)$ of edges in an n-vertex graph that does not contain a copy of H?

The extremal number $ex(n, H)$ is usually referred to the Turán number of the graph H.

By the classical Erdős-Stone-Simonovits theorem [13, 15], we have

$$ex(n, H) = \left(1 - \frac{1}{\chi(H) - 1} + o(1)\right)\binom{n}{2}.$$

where $\chi(H)$ is the chromatic number of H. Therefore, the order of $ex(n, H)$ is known, unless H is a bipartite graph. One of the major open problems in extremal graph theory is determining the correct order of magnitude for $ex(n, H)$. Here, we will focus on the Turán number of complete bipartite graphs $K_{s,t}$.

Kővari, Sós and Turán [24] proved that, for $s \geq t$,

$$ex(n, K_{t,s}) \leq \tfrac{1}{2}(s-1)^{1/t} n^{2-1/t} + \tfrac{1}{2}(t-1)n. \tag{4.1}$$

The best known general lower bounds, obtained by probabilistic constructions, are

$$ex(n, K_{t,s}) = \Omega(n^{2-(s+t-2)/(st-1)}),$$

see Erdős and Spencer [14], and

$$ex(n, K_{t,t}) = \Omega((\log n)^{1/(t^2-1)} n^{2-(2/(t+1))}),$$

see Bohman and Keevash [8].

The upper bound (4.1) is conjectured to give the right order of magnitude for $ex(n, K_{t,s})$.

By the construction introduced by Kollár et al. [23] and later improved by Alon et al. [1], if $s \geq (t-1)! + 1$, then

$$ex(n, K_{t,s}) \geq \tfrac{1}{2}(1 + o(1))d_t(s-1)^{1/t}n^{2-1/t},$$

where d_t is some constant. Hence for $s \geq (t-1)! + 1$ the bound (4.1) is indeed sharp.

The graph introduced in [23] and [1] is usually referred to as the *Norm Graph* $N(q,t)$ and it is defined as follows. Let q be a prime power and let $\mathbb{F}_q^*$ be $\mathbb{F}_q \setminus \{0\}$. The vertex set of $N(q,t)$ is $\mathbb{F}_{q^{t-1}} \times \mathbb{F}_q^*$ and two vertices $(a,\alpha), (b,\beta) \in \mathbb{F}_{q^{t-1}} \times \mathbb{F}_q^*$ are adjacent if and only if $N_{t-1}(a+b) = \alpha\beta$, where $N_{t-1} : x \in \mathbb{F}_{q^{t-1}} \mapsto x^{1+q+\cdots q^{t-2}} \in \mathbb{F}_q$ is the usual norm function. In [23] and [1], it is shown that any system of t equations $N_{t-1}(x+a_i) = y\alpha_i$, with $(a_i, \alpha_i) \in \mathbb{F}_{q^{t-1}} \times \mathbb{F}_q^*$, $i = 1, 2, \ldots, t$ has at most $(t-1)!$ solutions $(x,y) \in \mathbb{F}_{q^{t-1}} \times \mathbb{F}_q^*$. Therefore, any t vertices of $N(q,t)$ have at most $(t-1)!$ common neighbours.

Our aim is to highlight the beautiful geometric structure behind the Norm Graph. Let $T := \{0, 1, \ldots, t-1\}$, then

$$N(a+b) = \sum_{S \subseteq T} \prod_{i \in S,\ j \notin S} a^{q^i} b^{q^j}.$$

Let $\mathcal{V}_{t-1} := \mathcal{V}_{t-1,\sigma}$ be the SLP-variety consisting of the tensors:

$$\{v \otimes v^{\sigma} \otimes \cdots \otimes v^{\sigma^{t-1}}, \langle v \rangle \in \mathrm{PG}(1, q^t)\},$$

hence $\mathcal{V}_{t-1} \subset \mathrm{PG}(2^{t-1} - 1, q)$ and we have

$$(1,x)^{v_{t,\sigma}} = \left(\prod_{i \in S} x^{q^i} \right)_{S \subseteq T},$$

$$(0,1)^{v_{t,\sigma}} = (0, 0, \ldots, 0, 1).$$

Let us fix an ordering for the family of subsets of T $\mathcal{P}(T)$, so that the i-th component of $(1,x)^{v_{t,\sigma}}$ is $\prod_{j \in S} x^{q^j}$ where S is the i-th element of $\mathcal{P}(T)$ and such that the $(2^{t-1} - i)$-th element $\mod 2^{t-1}$ is the complement of S.

Let $\mathcal{V}_{t-1}^* := \mathcal{V}_{t-1} \setminus (0,1)^{v_{t,\sigma}}$ and embed the $\mathrm{PG}(2^{t-1} - 1, q)$ containing $\mathcal{V}_{t-1}^*$ as the hyperplane H of $\mathrm{PG}(2^{t-1}, q)$ of equation $x_{2^{t-1}} = 0$.

Let $V = \mathbb{F}_{q^t}^{2^{t-1}}$ and $\beta : V \times V \to \mathbb{F}_{q^t}$ the bilinear form such that

$$\beta(\mathbf{x}, \mathbf{y}) = \sum_{i=0}^{2^{t-1}-1} x_i y_{2^{t-1}-i-1} + x_{2^{t-1}} y_{2^{t-1}}.$$

Let $\mathcal{V}_{t-1}^* P$ be the cone of $\mathrm{PG}(2^{t-1}, q)$ with base the $\mathcal{V}_{t-1}^*$ contained in H and vertex the point $P = (0, 0, \ldots, 0, 1) \in \mathrm{PG}(2^{t-1}, q)$. Finally, let $\perp$ be the polarity induced by β. We observe that $P^\perp = H$.

So now we can give this new description of the Norm Graph. The vertex set is $\mathcal{V}_{t-1}^* P \setminus \{\mathcal{V}_{t-1}^*, P\}$, and two points P_1, P_2 are adjacent vertices if and only if $P_2 \in P_1^\perp$.

Let $\{P_1, P_2, \ldots, P_t\}$ be t distinct vertices, then the common neighbors of $\{P_1, P_2, \ldots, P_t\}$ are the points of $\langle P_1, P_2, \ldots, P_t\rangle^\perp \cap (\mathcal{V}_{t-1}^* P \setminus \{\mathcal{V}_{t-1}^*, P\})$. If $P \in \langle P_1, P_2, \ldots, P_t\rangle$, then $\langle P_1, P_2, \ldots, P_t\rangle^\perp \subset H$ and since H does not contain vertices of the graph, those t points do not have any common neighbors. If If $P \notin \langle P_1, P_2, \ldots, P_t\rangle$, then the t points are projected onto t distinct points of $\mathcal{V}_{t-1}^*$, hence onto t linearly independent points. Therefore $\dim\langle P_1, P_2, \ldots, P_t\rangle = t - 1$ and $\dim\langle P_1, P_2, \ldots, P_t\rangle^\perp = 2^{t-1} - t$. Since $P \notin \langle P_1, P_2, \ldots, P_t\rangle^\perp$, $\langle P_1, P_2, \ldots, P_t\rangle^\perp$ is projected on a subspace of H of codimension $t - 1$. The variety $\mathcal{V}_{t-1}$ is a subvariety of the Segre variety $\Sigma_{2,t-1}$. It is well known that $\Sigma_{2,t-1}$ has dimension $t - 1$ and degree $(t - 1)!$, hence the general subspace of $\mathrm{PG}(2^{t-1} - 1, \overline{\mathbb{F}_q})$ of codimension $t - 1$ intersects $\Sigma_{2,t-1}$ in at most $(t - 1)!$ points. In [23], it is basically proved that $\langle P_1, P_2, \ldots, P_t\rangle^\perp$ is projected in a general subspace of H. Therefore, $\{P_1, P_2, \ldots, P_t\}$ have at most $(t-1)!$ common neighbors.

This geometric point of view has led to the following result.

Theorem 4.1. *[5] For $q > (t - 1)!$ the graph $N(q, t)$ contains no $K_{t+1,(t-1)!-1}$.*

In particular, when $t = 4$, we get

$$ex(n, K_{5,5}) \geq \tfrac{1}{2}(1 + o(1))n^{7/4}$$

which is an asymptotic improvement to the lower bounds of $ex(n, K_{5,s})$, $5 \leq s$. Also, in this case, the proof is very simple and it only uses the characterization of $t + 1$ linearly dependent points of $\mathcal{V}_{t-1}$.

Let $\{P_1, P_2, \ldots, P_5\}$ be 5 distinct points of $\mathcal{V}_3^* P \setminus \{\mathcal{V}_3^*, P\} \subset \mathrm{PG}(8, \mathbb{F}_q)$. Here we use the fact that $t+1$ points of $\mathcal{V}_{t-1}$ are either independent or they are in normal rational curve of $\mathcal{V}_{t-1}$, hence a 3-dimensional subspace intersects $\mathcal{V}_3$ in at most 4 points or in a normal rational cubic. If $P \in \langle P_1, P_2, \ldots, P_5\rangle$, as before, the points do not have any common neighbor. Let $P \notin \langle P_1, P_2, \ldots, P_5\rangle$, so they are projected into 5 distinct points of $\mathcal{V}_3^*$. Let $\dim\langle P_1, P_2, \ldots, P_5\rangle = 4$, then $\dim\langle P_1, P_2, \ldots, P_5\rangle^\perp = 3$. If $\langle P_1, P_2, \ldots, P_5\rangle^\perp$ is projected on a space intersecting $\mathcal{V}_3$ in a normal rational

cubic, then we would get a $K_{5,q+1}$, but $N(q,4)$ has no $K_{4,7}$, clearly a contradiction. Hence $\langle P_1, P_2, \ldots, P_5\rangle^\perp$ is projected on a space intersecting $\mathcal{V}_3$ in at most 4 points, so $\{P_1, P_2, \ldots, P_5\}$ have at most 4 common points. Suppose now that $\dim\langle P_1, P_2, \ldots, P_5\rangle = 3$, then $\langle P_1, P_2, \ldots, P_5\rangle$ is projected on a space intersecting $\mathcal{V}_3$ in a normal rational cubic. Again, since $N(q,4)$ has no $K_{4,7}$, $\langle P_1, P_2, \ldots, P_5\rangle^\perp$ must contain at most 4 points.

We wish to conclude this Section with some comments. The approach here has been the following: take an algebraic variety X of dimension $t-1$, embed X as a hyperplane section of a cone XP and a polarity $\perp$ such that $P^\perp$ is the hyperplane containing X. Then consider the graph G whose vertices are the points of $XP \setminus \{X, P\}$ and such that P_1 and P_2 are adjacent if and only if $P_2 \in P_1^\perp$. An algebraic variety defined over $\mathbb{F}_q$ of dimension $t-1$ has $\Theta(q^{t-1})$ points (see [25]); therefore G has $n = \Theta(q^t)$ vertices and degree $\Theta(q^{t-1})$. If any t points of X are linearly independent and $\deg(X) = d$, then G is a graph with roughly $n^{2-1/t}$ edges with no copies of $K_{t,d+1}$. Hence, with $\deg(X) = d < (t-1)!$, we could prove that the bound (4.1) is tight for bipartite graphs not included in the Alon, Rónyai and Szabó result. In particular, if we find a set X of points of PG$(2t-3,q)$, not necessarily an algebraic variety, such that $|X| = \Theta(q^{t-1})$ and such that any t points of X are linearly independent, then the bound (4.1) would be proven to be tight for any $K_{t,s}, s \geq t$, as $\dim\langle v_i, i = 1,2,\ldots,t\rangle = t-1$ and $\dim\langle v_i, i = 1,2,\ldots,t\rangle^\perp = 2t-2-t = t-2$ for any t vertices of G, and a $(t-2)$-subspace can contain at most $t-1$ points of X. For $t=3$, we can take $X = \mathcal{V}_2$ that turns out to be an elliptic quadric of PG$(3,q)$. An elliptic quadric is a quadric of PG$(3,q)$ not containing lines; therefore any line intersects the quadric in at most two points. In other words, $\mathcal{V}_2$ is a set of q^2+1 points of PG$(3,q)$ such that any 3 points are linearly independent. For $t \geq 4$, sets with the aforementioned properties are not known at the moment.

4.1 An Open Problem

The *Ramsey number* $R(s,t)$ is the smallest n such that every graph on n vertices either has a clique of size s or an independent set of size t. We recall that a *clique* of a graph is a set of pairwise adjacent vertices and an *independent set* is a set of pairwise non-adjacent vertices.

The best known lower bound on $R(2r+2, 2r+2)$ is

$$(1-o(1))\frac{\sqrt{2}}{e}(2r+2)2^{r+1},$$

obtained using the probabilistic method, and there is no known explicit construction that has exponential size.

Let β be an alternating bilinear form of $\mathbb{F}_2^{2r}$, and let G_r be the polarity graph

associated to it, that is, two non-zero vectors $\mathbf{x}$ and $\mathbf{y}$ of $\mathbb{F}_2^{2r}$ are adjacent if and only if $\beta(\mathbf{x}, \mathbf{y}) = 0$.

Lemma 4.2. *An independent set of G_r has size at most $2r + 1$.*

Proof. Let $S = \{v_1, v_2, \ldots, v_t\}$ be a independent set of size $t = 2r+2$. Since $t > 2r$, the set is linearly dependent, there exist $\lambda_i \in \mathbb{F}_2$ not all zero such that $\sum_{i=1}^{t} \lambda_i v_i = \mathbf{0}$. Hence, for each $j \in \{1, 2, \ldots, t\}$,

$$0 = \beta\left(\sum_{i=1}^{t} \lambda_i v_i, v_j\right) = \sum_{i \neq j} \lambda_i.$$

Let I and J be the identity and the all-one matrix of order t respectively. Hence, there exists a non-zero vector $(\lambda_1, \ldots, \lambda_t)$ in the null space of the matrix $J - I$. But the eigenvalues of $J - I$ are $t - 1$ and -1, both of which are non-zero in $\mathbb{F}_2$, a contradiction. □

Hence G_r could be a good candidate to get a deterministic lower bound for $R(2r + 2, 2r + 2)$. We need to answer the following question:

What is the largest number of points in a symplectic polar space of rank r over $\mathbb{F}_2$, with the property that every generator meets this set of points in at most $2r + 1$ points?

In other words, what is the best construction for a partial $(2r + 1)$-ovoid? By counting, we see that such a set has cardinality at most $(2r + 1)(2^r + 1)$.

We stress out that any construction of size c^r, for some constant c, meeting each generator in at most ar points, for some fixed constant a, would be a huge breakthrough.

An approach could be the following: trying to construct an algebraic variety $\mathcal{V}$ of $\mathrm{PG}(2r - 1, 2)$ of dimension r such that the generators of a fixed symplectic polarity are in general position with respect to it, that is each generator meets $\mathcal{V}$ in at most d points, with d being the degree of $\mathcal{V}$. The problem here is that what happens over $\mathbb{F}_2$ does not say much about the behaviour of a variety over $\overline{\mathbb{F}_2}$, that is, we need to define $\mathcal{V}$ over $\mathbb{F}_2$, but, as we have already mentioned, the notions of dimension and degree need to be studied on $\mathcal{V}_{\overline{\mathbb{F}_2}}$.

Acknowledgment I wish to thank Anurag Bishnoi for introducing me to this beautiful problem.

References

[1] N. Alon, L. Rónyai, and T. Szabó. Norm-graphs: variations and applications. *J. Combin. Theory Ser. B*, 76(2):280–290, 1999.

[2] S. Ball. On sets of vectors of a finite vector space in which every subset of basis size is a basis. *J. Eur. Math. Soc. (JEMS)*, 14(3):733–748, 2012.

[3] S. Ball and M. Lavrauw. Arcs in finite projective spaces. *EMS Surv. Math. Sci.*, 6(1-2):133–172, 2019.

[4] S. Ball and V. Pepe. Asymptotic improvements to the lower bound of certain bipartite Turán numbers. *Combin. Probab. Comput.*, 21(3):323–329, 2012.

[5] S. Ball and V. Pepe. Forbidden subgraphs in the norm graph. *Discrete Math.*, 339(4):1206–1211, 2016.

[6] E. R. Berlekamp. *Algebraic coding theory*. McGraw-Hill Book Co., New York-Toronto-London, 1968.

[7] A. Bernal. On the existence of absolutely maximally entangled states of minimal support. *Quant. Phys. Lett.*, 6:1–3, 2017.

[8] T. Bohman and P. Keevash. The early evolution of the H-free process. *Invent. Math.*, 181(2):291–336, 2010.

[9] R. C. Bose. On some connections between the design of experiments and information theory. *Bull. Inst. Internat. Statist.*, 38:257–271, 1961.

[10] E. Crestani and A. Lucchini. d-Wise generation of prosolvable groups. *J. Algebra*, 369:59–69, 2012.

[11] M. A. de Boer. Almost MDS codes. *Des. Codes Cryptogr.*, 9(2):143–155, 1996.

[12] N. Durante, G. Longobardi, and V. Pepe. (d, σ)-Veronese variety and some applications. *Des. Codes Cryptogr.*, 91(5):1911–1921, 2023.

[13] P. Erdős and M. Simonovits. A limit theorem in graph theory. *Studia Sci. Math. Hungar.*, 1:51–57, 1966.

[14] P. Erdős and J. Spencer. *Probabilistic methods in combinatorics*. Probability and Mathematical Statistics, Vol. 17. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1974.

[15] P. Erdős and A. H. Stone. On the structure of linear graphs. *Bull. Amer. Math. Soc.*, 52:1087–1091, 1946.

[16] O. Farràs, C. Padró, C. Xing, and A. Yang. Natural generalizations of threshold secret sharing. *IEEE Trans. Inform. Theory*, 60(3):1652–1664, 2014.

[17] L. Giuzzi and V. Pepe. Families of twisted tensor product codes. *Des. Codes Cryptogr.*, 67(3):375–384, 2013.

[18] L. Giuzzi and V. Pepe. On some subvarieties of the Grassmann variety. *Linear Multilinear Algebra*, 63(11):2121–2134, 2015.

[19] J. Harris. *Algebraic geometry*, volume 133 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1992. A first course.

[20] J. W. P. Hirschfeld and L. Storme. The packing problem in statistics, coding theory and finite projective spaces: update 2001. In *Finite geometries*, volume 3 of *Dev. Math.*, pages 201–246. Kluwer Acad. Publ., Dordrecht, 2001.

[21] J. W. P. Hirschfeld and J. A. Thas. *General Galois geometries*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 1991. Oxford Science Publications.

[22] H. Kaneta and T. Maruta. An elementary proof and an extension of Thas' theorem on k-arcs. *Math. Proc. Cambridge Philos. Soc.*, 105(3):459–462, 1989.

[23] J. Kollár, L. Rónyai, and T. Szabó. Norm-graphs and bipartite Turán numbers. *Combinatorica*, 16(3):399–406, 1996.

[24] T. Kövari, V. T. Sós, and P. Turán. On a problem of K. Zarankiewicz. *Colloq. Math.*, 3:50–57, 1954.

[25] S. Lang and A. Weil. Number of points of varieties in finite fields. *Amer. J. Math.*, 76:819–827, 1954.

[26] G. Lunardon. Normal spreads. *Geom. Dedicata*, 75(3):245–261, 1999.

[27] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. I*. North-Holland Mathematical Library, Vol. 16. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977.

[28] V. Pepe. On the algebraic variety $\mathcal{V}_{r,t}$. *Finite Fields Appl.*, 17(4):343–349, 2011.

[29] B. Segre. Curve razionali normali e k-archi negli spazi finiti. *Ann. Mat. Pura Appl. (4)*, 39:357–379, 1955.

[30] B. Segre. Ovals in a finite projective plane. *Canadian J. Math.*, 7:414–416, 1955.

[31] B. Segre. Ovali e curve σ nei piani di Galois di caratteristica due. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat. (8)*, 32:785–790, 1962.

[32] B. Segre. Teoria di Galois, fibrazioni proiettive e geometrie non desarguesiane. *Ann. Mat. Pura Appl. (4)*, 64:1–76, 1964.

[33] B. Segre. Introduction to Galois geometries. *Atti Accad. Naz. Lincei Mem. Cl. Sci. Fis. Mat. Natur. Sez. Ia (8)*, 8:133–236, 1967.

[34] J. A. Thas. Normal rational curves and k-arcs in Galois spaces. *Rend. Mat. (6)*, 1:331–334, 1968.

Part IV

The Geometry of Semifields

John Sheekey

University College Dublin,
School of Mathematics and Statistics,
Science Centre - East Belfield Dublin 4,
Ireland

email: john.sheekey@ucd.ie

Contents

1 Motivating Examples **101**
1.1 Affine and Projective Planes 101
1.2 Spreads 102
1.3 Rank-Metric Codes 102
1.4 Tensors 103
1.5 The Answer is Always Semifields 103

2 Semifields **105**
2.1 Beginnings 105
2.2 Structure and Equivalence 110
2.3 Spread Sets 112
2.4 Knuth Orbit 116
2.5 Spreads and Translation Planes 118

3 Constructions, Classifications, and Invariants **121**
3.1 Representations and Constructions 121
3.2 Biprojective Polynomials 125
3.3 Cyclic Semifields and Skew Polynomial Rings 126
3.4 Classifications 129
3.5 Other Topics 131

4 The Geometry of Matrices and Tensors **133**
4.1 Matrices and Linear Maps 133
4.2 Linear Sets and Semifields 136
4.4 The Tensor Rank of a Semifield 140

Bibliography **143**

Chapter 1

Motivating Examples

1.1 Affine and Projective Planes

Consider the standard two-dimensional *Euclidean plane*, where *points* are regarded to be vectors in $\mathbb{R}^2 = \mathrm{V}(2, \mathbb{R})$ and *lines* are straight lines. We can regard the straight lines equivalently as solutions sets of linear equations

$$\{(x, y) : x, y \in \mathbb{R}, ax + by + c = 0\},$$

or in parametric form as either

$$\{(x_0, y_0) + \lambda(x_1, y_1) : \lambda \in \mathbb{R}\} = u + \langle v \rangle_{\mathbb{F}_q}$$

or

$$\{(x, mx + c) : x \in \mathbb{R}\}; \qquad \{(0, x) : x \in \mathbb{R}\}.$$

We all know that two different straight lines either meet in one point or don't meet at all, and they don't meet if and only if they have the same *slope* m. We call the lines that don't meet *parallel*. How do we prove this?

Consider $\{(x, mx + c) : x \in \mathbb{R}\} \cap \{(x, m'x + c') : x \in \mathbb{R}\}$. Then we must have $mx + c = m'x + c'$, and rearranging we get $(m - m')x = (c - c')$, which has a unique solution if $m - m' \neq 0$, and no solution if $m - m' = 0, c - c' \neq 0$.

This set of points and lines gives an *affine plane*, denoted by $\mathrm{AG}(2, \mathbb{R})$; indeed, it is the motivating example to study more general objects with the same properties. Adding *points at infinity* indexed by the slopes, we can extend this to a *projective plane*, denoted by $\mathrm{PG}(2, \mathbb{R})$.

Definition 1.1. An *affine plane* is a point-line incidence geometry satisfying the following axioms:

- Any two distinct points lie on a unique line.

- Given any line and any point not on that line there is a unique line which contains the point and does not meet the given line.
- There exist four points such that no three are collinear.

Definition 1.2. A *projective plane* is a point-line incidence geometry satisfying the following axioms:

- Any two distinct points lie on a unique line.
- Any two distinct lines meet in a unique point.
- There exist four points such that no three are collinear.

Clearly the field $\mathbb{R}$ is not special here; all we needed was the fact that $\mathbb{R}$ is a field. Hence for any field $\mathbb{F}$ we can define affine and projective planes $\mathrm{AG}(2, \mathbb{F})$ and $\mathrm{PG}(2, \mathbb{F})$ respectively. When $\mathbb{F} = \mathbb{F}_q$ is a finite field, we denote the planes by $\mathrm{AG}(2, q)$ and $\mathrm{PG}(2, q)$.

However, you may have noticed that we did not in fact use all of the axioms of a field in order to determine the sizes of the intersection of two lines. **Which ones did we not need?**

1.2 Spreads

When studying finite geometry, you may have come across the concept of a *spread*; a partition of the points of a projective space into pairwise disjoint projective subspaces of a fixed dimension, or equivalently a partition of the nonzero vectors of a vector space into vector subspaces of a fixed dimension.

What are the nicest spreads?

1.3 Rank-Metric Codes

A *linear code* is a subspace of a vector space endowed with some metric. The first metric one usually encounters is the *Hamming metric*, where the distance between two vectors is defined as the number of positions in which they differ. Such codes have been studied and used in applications for many years.

More recently, a different framework has emerged as an interesting and useful alternative, namely *rank-metric codes*. Here codes are subspaces of matrices, and the distance between two matrices is defined as the *rank* of their difference.

Can we construct or classify classes of rank-metric codes with certain properties

1.4 Tensors

Tensors can be thought of as higher-dimensional analogues of linear maps and matrices; for example, a 3-tensor can be represented by a 3-dimensional array of elements from a field.

A square matrix is *nonsingular* if no nontrivial linear combination of its columns is zero, i.e. its columns are linearly independent. A cubical array is nonsingular if no nontrivial linear combination of its *slices* is a singular matrix.

Can we count and/or characterise the nonsingular 3-tensors?

1.5 The Answer is Always Semifields

Always.

Chapter 2

Semifields

Finite fields underpin a large portion of discrete mathematics and incidence geometry. They can be used to construct classical examples of objects such as *affine/projective planes/spaces*. However, many applications of finite fields do not require all the axioms of a field. In their place, algebraic structures known as *semifields* can be used. These objects are interesting for a variety of reasons, from their own algebraic structure, to connections with projective planes, spreads, rank-metric codes, tensors, and many other areas of finite geometry. Indeed, they pop up unexpectedly in surprising places, so if you are familiar with them they can often be a source of examples, results, and techniques for whatever geometric object you are studying.

2.1 Beginnings

Let us start at the beginning; although we assume that the reader is familiar with finite fields already, it will be useful to recap the basics for the purposes of comparing and contrasting once we move on to semifields.

Definition 2.1. A *field* is a set $\mathbb{F}$ with two binary operations, addition and multiplication, with two distinct elements 0 and 1, satisfying the following identities for all $a, b, c \in \mathbb{F}$:

Addition	Commutative	$a + b = b + a$
	Associative	$(a + b) + c = a + (b + c)$
	Identity	$0 + a = a + 0 = a$
	Inverses	$\exists! x$ such that $a + x = b$
Multiplication	Commutative	$ab = ba$
	Associative	$(ab)c = a(bc)$
	Identity	$1a = a1 = a$
	Inverses	$\exists! x$ such that $ax = b \quad (a \neq 0)$
	Distributive	$a(b + c) = ab + ac$
		$(a + b)c = ac + bc$

Finite fields are sometimes referred to as *Galois fields*, since the study of finite field theory in its modern sense originates in many ways from the work of Galois in the 1830s (although fields of prime order were implicitly studied much earlier, in particular by Gauss). Galois showed that there exists a field of each prime power order, and also proved many results about the structure of the examples he constructed.

Theorem 2.2 (Galois 1830)**.** *Let q be a power of a prime p with $p = q^h$. Then there exists an irreducible polynomial $f(x)$ of degree h in $\mathbb{F}_p[x]$. Moreover, the quotient $\mathbb{F}_p[x]/(f(x))$ is a finite field containing q elements.*

The full classification of finite fields was established by Moore.

Theorem 2.3 (Moore (1893))**.** *Every finite field has prime power order, and there is a unique field (up to isomorphism) of each prime power order.*

We will usually work with a field of order q (which may or may not be prime), and a field containing it of order q^n, which we denote as usual by $\mathbb{F}_q$ and $\mathbb{F}_{q^n}$ respectively.

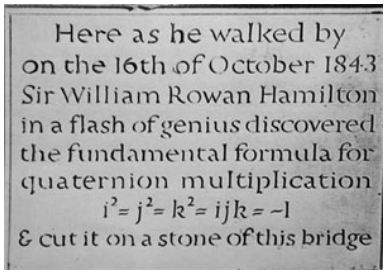
Proposition 2.4. *Consider a field $\mathbb{F}_{q^n}$ containing $\mathbb{F}_q$.*

- *$(\mathbb{F}_{q^n}, +)$ has the structure of a vector space over $\mathbb{F}_q$.*
- *Every element x of $\mathbb{F}_{q^n}$ satisfies the equation $x^{q^n} - x = 0$.*
- *The multiplicative group $(\mathbb{F}_{q^n}^\times, \cdot)$ is cyclic.*
- *The group of field automorphisms of $\mathbb{F}_{q^n}$ fixing each element of $\mathbb{F}_q$ is cyclic of order n, and is generated by the Frobenius automorphism $x \mapsto x^q$.*
- *The trace map $x \mapsto \mathrm{tr}(x) := x + x^q + \cdots + x^{q^{n-1}}$ is a q^{n-1}-to-one map from $\mathbb{F}_{q^n}$ to $\mathbb{F}_q$. It is a surjective group homomorphism from the additive group of $\mathbb{F}_{q^n}$ to the additive group of $\mathbb{F}_q$.*

- *The norm map $x \mapsto N(x) := x^{1+q+\cdots+q^{n-1}}$ is a $\frac{q^n-1}{q-1}$-to-one map from $\mathbb{F}_{q^n}^\times$ to $\mathbb{F}_q^\times$. It is a surjective group homomorphism from the multiplicative group of $\mathbb{F}_{q^n}$ to the multiplicative group of $\mathbb{F}_q$.*

- *There exists a normal basis, i.e. there exists an element $\alpha \in \mathbb{F}_{q^n}$ such that $\{\alpha, \alpha^q, \ldots, \alpha^{q^{n-1}}\}$ is an $\mathbb{F}_q$-basis for $\mathbb{F}_{q^n}$.*

Famously, Hamilton demonstrated the existence of a non-commutative analogue of a field, by defining a multiplication on a four-dimensional real vector space. The resulting algebra is known as the *quaternions*.



Definition 2.5. An *(associative) division algebra* (or *skew field*) is a set $\mathbb{F}$ with two binary operations, addition and multiplication, with two distinct elements 0 and 1, satisfying the following identities for all $a, b, c \in \mathbb{F}$:

Addition	Commutative	$a + b = b + a$
	Associative	$(a + b) + c = a + (b + c)$
	Identity	$0 + a = a + 0 = a$
	Inverses	$\exists! x$ such that $a + x = b$
Multiplication	~~Commutative~~	~~$ab = ba$~~
	Associative	$(ab)c = a(bc)$
	Identity	$1a = a1 = a$
	Inverses	$\exists! x$ such that $ax = b \quad (a \neq 0)$
		$\exists! y$ such that $ya = b \quad (a \neq 0)$
	Distributive	$a(b + c) = ab + ac$
		$(a + b)c = ac + bc$

The additive structure forms a vector space over some field.

For finite algebras, a famous result often referred to as *Wedderburn's Little Theorem*, shows that we do not obtain any new structures. Note however that Wedderburn's work contained errors that were later fixed by Dickson.

Theorem 2.6 (Wedderburn-Dickson Theorem)**.** *Every finite associative division algebra is a finite field.*

Definition 2.7. A *semifield* is a set $\mathbb{S}$ with two binary operations, addition and multiplication, with two distinct elements 0 and 1, satisfying the following identities for all $a, b, c \in \mathbb{S}$:

Addition	Commutative	$a + b = b + a$
	Associative	$(a + b) + c = a + (b + c)$
	Identity	$0 + a = a + 0 = a$
	Inverses	$\exists! x$ such that $a + x = b$
Multiplication	~~Commutative~~	~~$ab = ba$~~
	~~Associative~~	~~$(ab)c = a(bc)$~~
	Identity	$1a = a1 = a$
	Inverses	$\exists! x$ such that $ax = b \quad (a \neq 0)$
		$\exists! y$ such that $ya = b \quad (a \neq 0)$
	Distributive	$a(b + c) = ab + ac$
		$(a + b)c = ac + bc$

Exercise 2.8. Show that in a *finite* semifield, the non-existence of non-trivial zero divisors implies that linear equations have a unique solution; that is, for any $a, b \in \mathbb{S}$ with $a \neq 0$, there exists a unique x such that $ax = b$, and a unique y such that $ya = b$.

Exercise 2.9. Show that $(\mathbb{S}, +)$ can be viewed as a vector space over some field (that is, an elementary abelian p-group for some prime p).

We could also relax our axioms in different ways. The following are algebraic structures where

- addition forms a vector space;
- multiplication is right-distributive over addition;
- left- and right-multiplication by every nonzero element defines a permutation.

	Comm.	Assoc.	Left-distr.
Field	✓	✓	✓
Division ring		✓	✓
Semifield			✓
Nearfield		✓	
Quasifield			

Note that we can have a semifield which is commutative but not associative. We may sometimes have reason to not assume a multiplicative identity; in this case the structure is called a *presemifield*. However we will see later that every presemifield is equivalent (*isotopic*) to a semifield.

Definition 2.10. A *presemifield* is a set $\mathbb{S}$ with two binary operations, addition and multiplication, with a distinct element 0, satisfying the following identities for all $a, b, c \in \mathbb{S}$:

Addition	Commutative	$a + b = b + a$
	Associative	$(a + b) + c = a + (b + c)$
	Identity	$0 + a = a + 0 = a$
	Inverses	$\exists! x$ such that $a + x = b$
Multiplication	~~Commutative~~	~~$ab = ba$~~
	~~Associative~~	~~$(ab)c = a(bc)$~~
	~~Identity~~	~~$1a = a1 = a$~~
	Inverses	$\exists! x$ such that $ax = b \quad (a \neq 0)$
		$\exists! y$ such that $ya = b \quad (a \neq 0)$
	Distributive	$a(b + c) = ab + ac$
		$(a + b)c = ac + bc$

2.1.1 First Examples

The first non-trivial example of a finite semifield was constructed by Dickson in 1905 [12], thus establishing that the Wedderburn-Dickson theorem does not extend to semifields. By non-trivial we mean not equivalent to a field; we will define precisely what we mean later.

Many of the known examples of semifields are constructed by starting with a finite field and modifying the multiplication in some way.

Exercise 2.11. Let $q = p^h$ be a power of an odd prime p with $h > 1$. Show that there exists an element $k \in \mathbb{F}_q$ such that $x^2 - k$ is irreducible in $\mathbb{F}_q[x]$. Show that the multiplication in the finite field $\mathbb{F}_q[x]/(x^2 - k)$ can be written as

$$(a + bx)(c + dx) = (ac + kbd) + (ad + bc)x.$$

Exercise 2.12. Show that the following multiplications possess non-trivial zero divisors.

$$(a + bx) \star_1 (c + dx) = (ac + kb^p d) + (ad + bc)x;$$
$$(a + bx) \star_2 (c + dx) = (ac + k(bd)^p) + (ad + bc)x$$

This is one of the semifields constructed by Dickson. We will postpone the proof that this is not equivalent to a field.

Exercise 2.13. Let $\mathbb{S} = (\mathbb{F}_{q^2}, +, \star)$, where $(\mathbb{F}_{q^2}, +)$ is the usual additive group of $\mathbb{F}_{q^2}$, and $\star$ is defined by

$$x \star y = \begin{cases} xy & \text{if y is a square} \\ x^q y & \text{otherwise} \end{cases}$$

Show that this defines the multiplication of a quasifield, but not a semifield.

Exercise 2.14. Let $\mathbb{S} = (\mathbb{F}_{q^n}, +, \star)$, where $(\mathbb{F}_{q^n}, +)$ is the usual additive group of $\mathbb{F}_{q^n}$, and $\star$ is defined by

$$x \star y = xy - \eta x^{q^i} y^{q^j},$$

where η is a fixed element of $\mathbb{F}_{q^n}$. Determine conditions on η which ensure that $\star$ has no non-trivial zero divisors. Does this multiplication have an identity element?

These (pre)semifields are known as *Generalised Twisted Fields*, and are due to Albert (1961) [1].

2.2 Structure and Equivalence

2.2.1 Isotopy

It is natural to ask which operations preserve the property of being a semifield. For groups, rings, fields, we usually consider isomorphisms; this is necessary in order to preserve associativity. However when dealing with nonassociative structures, we can define a more general notion of equivalence.

Since all finite-dimensional vector spaces over a finite field are equivalent (via an invertible additive map), we will assume that all semifields of a fixed order have the same additive structure, with different multiplications.

Definition 2.15. Two presemifields $(\mathbb{S}_1, +, \star)$ and $(\mathbb{S}_2, +, \circ)$ are *isotopic* if there exist invertible additive maps A_1, A_2, A_3 such that

$$(x \star y)^{A_1} = x^{A_2} \circ y^{A_3}$$

for all $x, y \in \mathbb{S}_1$. The set of presemifields isotopic to $\mathbb{S}_1$ is called the *isotopy class* of $\mathbb{S}_1$, and denoted by $[\mathbb{S}_1]$.

Exercise 2.16. Convince yourself that isotopy defines an equivalence relation on presemifields.

Exercise 2.17. Let $\mathbb{S}$ be a presemifield, and u an arbitrary nonzero element of $\mathbb{S}$. Show that the multiplication $\circ$ defined by $x \star y = (x \star u) \circ (u \star y)$ is a semifield with multiplicative identity $u \star u$. Verify that these two multiplications are isotopic.

Exercise 2.18. Show that the two presemifields defined in Exercise 2.12 are isotopic.

Exercise 2.19. Calculate the multiplication of a semifield isotopic to the presemifield defined in Exercise 2.14.

Exercise 2.20. Calculate the multiplication of a semifield isotopic to the presemifield defined in Exercise 2.12.

2.2.2 Centre and Nuclei

Definition 2.21. Let $\mathbb{S}$ be a semifield. The *left, middle and right nucleus* are defined respectively as

$$N_\ell = \{a \in \mathbb{S} \mid (ab)c = a(bc)\ \forall b, c \in \mathbb{S}\}$$
$$N_m = \{b \in \mathbb{S} \mid (ab)c = a(bc)\ \forall a, c \in \mathbb{S}\}$$
$$N_r = \{c \in \mathbb{S} \mid (ab)c = a(bc)\ \forall a, b \in \mathbb{S}\}$$

The *nucleus* is defined as the intersection of these three sets.

Note that we define these for semifields, not presemifields. We need to be a bit careful when extending this definition to presemifields.

Definition 2.22. The *centre* $Z(\mathbb{S})$ is defined as the set of elements in the nucleus which commute with all elements of $\mathbb{S}$.

The centre is the largest field over which $\mathbb{S}$ is a division algebra. Note that in an associative ring the centre is usually the set of elements which commute with all others; in the nonassociative setting we refer to this as the *commutative centre*.

Exercise 2.23. Show that the nuclei are all division rings, and centre is a field.

Exercise 2.24. Suppose that $\mathbb{S}_1$ and $\mathbb{S}_2$ are isotopic semifields. Calculate the nuclei of one in terms of the nuclei of the other. Hence show that the orders of the nuclei are isotopy invariants.

Exercise 2.25. Show that $\mathbb{S}$ is a *left vector space* over its left nucleus.

2.2.3 Autotopism Group

Just as an associative structure has an automorphism group, a semifield has an *autotopism group*.

Definition 2.26. An *autotopism* of a semifield $(\mathbb{S}, +, \star)$ is a triple A_1, A_2, A_3 of invertible additive maps such that

$$(x \star y)^{A_1} = x^{A_2} \star y^{A_3}$$

for all $x, y \in \mathbb{S}$. The set of autotopisms together with the operation $(A_1, A_2, A_3)(B_1, B_2, B_3') = (A_1B_1, A_2B_2, A_3B_3)$ is called the *autotopism group* of $\mathbb{S}$, and is denoted by $\mathrm{Autt}(\mathbb{S})$.

We use the double "t" to distinguish from the usual automorphism group, which does make sense and is sometimes studied for semifields.

Exercise 2.27. Consider the generalised twisted fields introduced in Exercise 2.14. Suppose (A_1, A_2, A_3) is an autotopism, and suppose that $A_1, A_2, A_3 \in \Gamma\mathrm{L}(1, q^n)$; that is, $A_1(x) = \alpha x^{q^a}, A_2(x) = \beta x^{q^b}, A_3(x) = \gamma x^{q^c}$ for some $\alpha, \beta, \gamma \in \mathbb{F}_{q^n}^\times$ and some integers a, b, c. Determine conditions under which (A_1, A_2, A_3) is an autotopism, and count the number of autotopisms you obtain.

Exercise 2.28. Suppose $\mathbb{S}$ is a semifield, and $a \in N_\ell(\mathbb{S})$. Let L_a denote the map $x \mapsto a \star x$. Show that (L_a, L_a, I) is an autotopism of $\mathbb{S}$. Can you relate the elements of the other nuclei to elements of the autotopism group?

Remark 2.29. While studying semifields up to autotopism is the more common for those working in finite geometry, it is also interesting to study automorphism classes. Clearly isotopy classes are unions of isomorphism classes. It turns out that the number of isomorphism classes in an isotopy class is equal to the number of orbits of points in the corresponding projective plane under the collineation group of the plane. Liebler and Kallaher [28] (1979) conjectured that this number is at least five for any non-desarguesian plane (and proved it under certain assumptions). The general case was proved by Ganley and Jha. [20] (1986).

Remark 2.30. Not much is known about the possible structure of the autotopism group of a semifield. It has been conjectured that the autotopism group is *solvable*; however, there is not much evidence to support this conjecture, beyond the lack of examples. Recently it was shown in [32] that the simple (and hence not solvable) group A_5 cannot appear as a subgroup of the autotopism group. There remains much to be explored in this direction.

2.3 Spread Sets

As we have somewhat hinted at through the exercises, the multiplication in a semifield can be used to define additive maps in two ways.

Definition 2.31. Let $\mathbb{S}$ be a presemifield with multiplication $\star$. Then the maps

$$L_x : y \mapsto x \star y$$
$$R_y : x \mapsto x \star y$$

are known as the *maps of left-* and *right-multiplication* respectively.

Exercise 2.32. Verify that L_x and R_y define invertible additive maps on $\mathbb{S}$ whenever $x, y \neq 0$. If we replace $\mathbb{S}$ by a quasifield, are these maps still additive?

We have a choice to make here: do we work with left- or right-multiplication? It does not make much difference in the grand scheme of things, but we have to pick one and stick to it. After tossing a coin, we will stick to right-multiplication for the majority of these notes.

Definition 2.33. The *spread set* of a presemifield (or prequasifield) $\mathbb{S}$ is denoted by $C(\mathbb{S})$ and defined as

$$\{R_y : y \in \mathbb{S}\}.$$

Exercise 2.34. Show that $C(\mathbb{S})$ is additively closed if $\mathbb{S}$ is a presemifield. Show that moreover it is a vector space over the centre $Z(\mathbb{S})$.

Exercise 2.35. A *nearfield* is a quasifield in which multiplication is associative. Show that the spread set of of a nearfield is closed under multiplication (and so the nonzero elements form a subgroup).

Conversely, given a set of endomorphisms with the appropriate properties, we can define a prequasifield.

Lemma 2.36. *Let C be a subset of additive maps from a finite vector space V to itself such that the difference of any two distinct elements of C is invertible, and such that $|C| = |V|$. Let ϕ be any bijection from V to C. Then the multiplication $\star_\phi$ defined by*

$$x \star_\phi y := \phi(y)(x)$$

defines a prequasifield multiplication on V. If C is additively closed, and ϕ is additive, then $\star_\phi$ defines a presemifield.

Exercise 2.37. Show that a spread set containing the identity map defines a unique semifield.

Exercise 2.38. Suppose (A_1, A_2, A_3) is an isotopism from $\mathbb{S}_1 = (V, \star)$ to $\mathbb{S}_2 = (V, \circ)$. Show that

$$A_1 R_y^\star A_3 = R^\circ_{A_2(y)},$$

and hence $A_1 C(\mathbb{S}_1) A_3 = C(\mathbb{S}_2)$.

Definition 2.39. A *rank metric code* is a subset C of $\mathrm{Hom}_F(U, V)$ endowed with the distance function

$$d(X, Y) = \mathrm{rank}(X - Y),$$

where rank denotes the usual linear algebraic rank of a linear map. If C is additively closed (resp. linear) then the code is said to be *additive* (resp. *linear*).

It is natural to study codes up to *isometries*; that is, maps on the ambient space which preserve the distance. So we need to know which maps ϕ satisfy $\mathrm{rank}(\phi(X) - \phi(Y)) = \mathrm{rank}(X - Y)$ for all $X, Y \in \mathrm{Hom}_F(U, V)$.

Definition 2.40. Two additive rank metric codes C_1, C_2 are said to be *linearly equivalent* if there exist invertible maps $A_1 \in \mathrm{Hom}_F(U, U)$, $A_2 \in \mathrm{Hom}_F(V, V)$ such that $A_1 C_1 A_2 = C_2$.

Example 2.41. The following are the spread sets of the three isotopy classes of finite semifields of order 16.

A basis for $C(\mathbb{F}_{2^4})$ over $\mathbb{F}_2$ is

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

$$= \{1, M, M^2, M^3\},$$

where M is the companion matrix of an irreducible polynomial $x^4 + x + 1$ of degree four in $\mathbb{F}_2[x]$.
A basis for $C(\mathbb{S}_1)$ over $\mathbb{F}_2$ is

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

A basis for $C(\mathbb{S}_2)$ over $\mathbb{F}_2$ is

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

2.3.1 Nuclei, Idealisers, and Centralisers

Note how in the basis for $C(\mathbb{S}_1)$ in the previous section, each matrix can be formed from a block matrix composed of 2×2 blocks.

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

Moreover, the only blocks which appear are from the set

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

In fact, this set is $C(\mathbb{F}_4)$, and so we could think of $C(\mathbb{S}_1)$ as a $\mathbb{F}_2$-subspace of $M_2(\mathbb{F}_4)$, with $\mathbb{F}_2$-basis

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

We will see now that this occurs because the left nucleus is isomorphic to $\mathbb{F}_4$. However not every spread set for a semifield isotopic to $\mathbb{S}_1$ will be in such an easily recognised form. Given a spread set, how do we determine its nuclei? When can we embed it in a space of smaller matrices over a larger field?

Exercise 2.42. Show that if $c \in N_r(\mathbb{S})$, then $R_{b \circ c} = R_c R_b$. Hence show that $R_c C(\mathbb{S}) = C(\mathbb{S})$.

Here by $R_c R_b$ we mean the linear map $a \mapsto R_c(R_b(a))$. Note that all of this depends on conventions; not only whether we choose maps of left- or right-multiplication, but also whether we regard linear maps as acting on rows or columns. There is little consensus in the literature; the safest thing to do is to calculate it again each time!

Definition 2.43. The *left idealiser*, *right idealiser*, and *centraliser* of a subspace C of linear maps (or matrices) are defined respectively as

$$\begin{aligned} \mathcal{I}_\ell(C) &:= \{X : XC \subset C\} = \{X : XY \in C \; \forall Y \in C\} \\ \mathcal{I}_r(C) &:= \{X : CX \subset C\} = \{X : YX \in C \; \forall Y \in C\} \\ \mathrm{Cent}(C) &:= \{X : XY = YX \; \forall Y \in C\} \end{aligned}$$

It is straighforward to verify that each of these are subrings of the endomorphism ring.

Exercise 2.44. Calculate the idealisers of $C(\mathbb{S}_1)$.

From Exercise 2.42, we obtain an injection from $N_r(\mathbb{S})$ into $\mathcal{I}_\ell(C(\mathbb{S}))$. In fact we can show that this is a bijection. Similarly we can find bijections from the other nuclei into these sets defined purely in terms of the spread sets.

Lemma 2.45. *Let $\mathbb{S}$ be a semifield, and $C = C(\mathbb{S})$ its spread set. Then*

$$\begin{aligned} N_r(\mathbb{S}) &\simeq \mathcal{I}_\ell(C); \\ N_m(\mathbb{S}) &\simeq \mathcal{I}_r(C); \\ N_\ell(\mathbb{S}) &\simeq \mathrm{Cent}(C). \end{aligned}$$

Note that for the left nucleus, the bijection is from a to L_a, rather than R_a. Both the idealisers are subspaces of the spread set, whereas the centraliser is not necesarily so.

Note also that for presemifields, we need to be more careful. A full description of the nuclei in terms of spread sets can be found in [39].

The idealisers and centralisers are relatively quick to calculate computationally; indeed they can be solved via systems of linear equations. They can be trickier to calculate theoretically from a given construction, but they are not too bad. Certainly they are easier to compute than abstract isotopy testing; so, if you find a construction for semifields, the first thing to do is to calculate the nuclei to narrow down the list of known semifields to which it might be isotopic.

Suppose now that we have a semifield $\mathbb{S}$ of order $q^n = q^{ms}$, with left nucleus isomorphic to $\mathbb{F}_{q^s}$ (or containing a field isomorphic to $\mathbb{F}_{q^s}$). Then we have a subspace of $M_n(\mathbb{F}_q)$ isomorphic to $\mathbb{F}_{q^s}$ which commutes with every element of $C(\mathbb{S})$. Conversely, this means that $C(\mathbb{S})$ is contained in the centraliser of a copy of $\mathbb{F}_{q^s}$

in $M_n(\mathbb{F}_q)$. It can be shown that this centraliser is in fact a subring of $M_n(\mathbb{F}_q)$ isomorphic to $M_m(\mathbb{F}_{q^s})$. Thus we can represent the spread set as an $\mathbb{F}_q$-subspace of $M_m(\mathbb{F}_{q^s})$.

The following theorem allows us to determine equivalences between semifields with known nuclei.

Theorem 2.46. *Suppose $\mathbb{S}, \mathbb{S}'$ are two semifields m-dimensional over its left nucleus, with left nucleus isomoprhic to $\mathbb{F}_{q^s}$. Then $\mathbb{S}$ and $\mathbb{S}'$ are isotopic if and only if there exist $X, Y \in \mathrm{GL}(m, q^s)$ and $\rho \in \mathrm{Aut}(\mathbb{F}_{q^s} : \mathbb{F}_q)$ such that*

$$C(\mathbb{S}') = XC(\mathbb{S})^\rho Y.$$

The history of this theorem is a little difficult to trace. In [34] it is attributed to Maduram. A more recent and direct proof can be found in [16].

2.3.2 Commutative Semifields and PN Functions

Many of the results on semifields in recent years have been due to the connection between *commutative semifields* and *perfect nonlinear (PN)* functions. Although we will not focus on this topic in these lectures, we briefly mention it in order to highlight this motivation.

Definition 2.47. A function $f : \mathbb{F}_{q^n} \to \mathbb{F}_{q^n}$ is *pperfect nonlinear (PN)* (or *planar*) if for every nonzero $a \in \mathbb{F}_{q^n}$, the map

$$x \mapsto f(x + a) - f(x)$$

is bijective.

Exercise 2.48. Suppose $\mathbb{S}$ is a commutative semifield of odd order. Show that the function $f(x) := x \circ x$ defines a PN function.

Although commutative semifields are more difficult to find, in fact some of the largest families of semifields found to date are commutative semifields, somewhat counterintuitively.

2.4 Knuth Orbit

Donald Knuth is a computer scientist and mathematician known for many important contributions; amongst them is the creation of the TEX typesetting system, which we all now use to write our mathematical documents, and writing the famous book *The Art of Computer Programming*.

However, he started his career as a finite geometer; his PhD thesis *Finite Semifields and Projective Planes* in 1963 (and the accompanying paper [31]) was in many ways

responsible for the resurgence in the study of semifields at the time. He provided constructions, and performed a full computer classification of semifields of order 16. He also provided a new way to find new semifields from old. He originally did this using *hypercubes*; we will see later that this can perhaps more naturally be understood using the language of *tensors*. For now though we will proceed as Knuth did.

Let $\{e_1, \ldots, e_n\}$ be an $\mathbb{F}_q$-basis for $\mathbb{S}$. Then there exist unique elements T_{ijk} of $\mathbb{F}_q$ such that

$$e_i \circ e_j = \sum_{k=1}^{n} T_{ijk} e_k.$$

Definition 2.49. The three-dimensional array $T(\mathbb{S})$ with (i, j, k)-entry T_{ijk} is referred to as a *hypercube representing* $\mathbb{S}$.

Note that the matrix of multiplication L_{e_i} has (j, k)-entry T_{ijk}. Hence we can think of T as "stacking" the elements of a basis of $C(\mathbb{S})$;

$$[(T_{1jk})_{j,k}, \ldots, (T_{njk})_{j,k}]$$

Theorem 2.50. *Let $\tau \in S_3$. Then the hypercube T^τ defined as $(T^\tau)_{ijk} = T_{\tau(ijk)}$ is a hypercube representing a semifield if and only if T is a hypercube representing a semifield.*

Thus from one isotopy class $[\mathbb{S}]$ of semifields, we can obtain up to six isotopy classes in this way.

Definition 2.51. The *Knuth orbit* of a semifield $\mathbb{S}$ is the set of isotopy classes

$$\mathcal{K}(\mathbb{S}) := \{[\mathbb{S}]^\tau : \tau \in S_3\}.$$

Exercise 2.52. Show that the isotopy class $[\mathbb{S}]^{(12)}$ contains the *opposite* semifield to $\mathbb{S}$; that is, the algebra with multiplication $x \star y := y \circ x$.

Note that the translation plane defined by the opposite semifield is isomorphic to the dual of the translation plane of the original semifield.

Exercise 2.53. Show that the spread set of a representative in the isotopy class $[\mathbb{S}]^{(23)}$ is equal to the transpose of the spread set of $\mathbb{S}$.

For these reasons we usually refer to $[\mathbb{S}]^{(12)}$ as the *dual semifield*, and $[\mathbb{S}]^{(23)}$ as the *transpose semifield*.

The Knuth orbit does not have to have six distinct isotopy classes; for example if the semifield is commutative, then there will be at most three distinct isotopy classes in the Knuth orbit. The Knuth orbit of a field contains only one isotopy class.

2.5 Spreads and Translation Planes

Definition 2.54. Let $\mathcal{S}$ be a set of subspaces of $V = V(n.\mathbb{F})$. The translation structure $T(\mathcal{S})$ is the incidence geometry whose points are the vectors of $V(n, \mathbb{F})$, and whose blocks are translations of the elements of $\mathcal{S}$; that is $\mathcal{B} = \{u + U : u \in V, U \in \mathcal{S}\}$. The sets $\{u + U : u \in V\}$ are called the *parallel classes* of $T(\mathcal{S})$.

Exercise 2.55. Determine the possible sizes of the intersection of two blocks.

2.5.1 Spreads

Definition 2.56. A *k-spread* of a vector space $V(n, \mathbb{F})$ is a set $\mathcal{S}$ of subspaces of dimension k such that every nonzero vector is in precisely one element of $\mathcal{S}$.

Exercise 2.57. Show that $T(\mathcal{S})$ is an affine plane if and only if $\mathcal{S}$ is a spread.

Exercise 2.58. Show that a k-spread exists in $V(n, q)$ only if k divides n.

For the converse, suppose we have a tower of field extensions $\mathbb{F} \subset \mathbb{K} \subset \mathbb{L}$. with $[\mathbb{K} : \mathbb{F}] = k$, $[\mathbb{L} : \mathbb{F}] = n$ (which implies that k divides n). Then it is straightforward to verify that the following is a k-spread of $\mathbb{L}$, viewed as an n-dimensional vector space over $\mathbb{F}$.

$$\mathcal{D} = \{\{ax : a \in \mathbb{K}\} : x \in \mathbb{L}^\times\}.$$

Such a spread is called a *desarguesian spread*. This is because the translation structure $T(\mathcal{D})$ satisfies Desargues' theorem. In the finite case, such field extensions are unique, and so there is only one choice.

There are many other ways to construct a spread. The most relevant to us is n-spreads in $V(2n, q)$. Suppose that $\mathcal{S}$ is such a spread. Up to an invertible linear transformation, we can assume without loss of generality that $\mathcal{S}$ contains

$$S_\infty := \{(0, x) : x \in \mathbb{F}_q^n\}; \qquad S_0 := \{(x, 0) : x \in V(n, q)\}.$$

Then every other element of $\mathcal{S}$, being an n-dimensional vector space and thus the image of an injective map from $V(n, q)$ into $V(2n, q)$ and meeting S_∞ trivially, is of the form

$$S_A := \{(x, A(x)) : x \in V(n, q)\}$$

for some linear map A from $V(n,q)$ to itself. Since S_A must meet S_0 trivially, A must be an invertible linear map.

Let $C(\mathcal{S}) = \{A \in \mathrm{End}_{\mathbb{F}_q}(V(n,q)) : S_A \in \mathcal{S}\}$. Then since for all $A, B \in C(\mathcal{S})$ with $A \neq B$ we have $S_A \cap S_B = 0$, we get that $A - B$ is invertible. Since a n-spread in $V(2n,q)$ must have $q^n + 1$ elements, we have $|C(\mathcal{S})| = q^n$. Therefore $C(\mathcal{S})$ is an *MRD code*, or the spread set of a quasifield.

Conversely, given a quasifield we can define a spread:

$$\mathcal{S}(Q) := \{\{(x, x \circ y) : x \in Q\} : y \in Q^\times\} \cup \{S_\infty\}.$$

Hence we get a correspondence between n-spreads in $V(2n,q)$ and quasifields of dimension n over $\mathbb{F}_q$.

How can we recognise spreads arising from semifields amongst these spreads?

For an $\mathbb{F}_q$-linear map A, define $\phi_A(x,y) = (x, y + A(x))$. Then $\phi_A(S_B) = S_{A+B}$, and $\phi_A(S_\infty) = S_\infty$. Hence we can easily see one direction of the following theorem; the other direction is more involved.

Lemma 2.59. *Let $\mathbb{S}$ be a semifield spread. Then there exists a subgroup of $\mathrm{GL}(2m,q)$ fixing one element of $\mathcal{S}(\mathbb{S})$ and acting transitively on the remaining elements of $\mathcal{S}(\mathbb{S})$. Conversely, any spread with such a group action is equivalent to one arising from a semifield.*

For this reason we call such a spread a *semifield spread*. The distinguished element which is fixed by this group action is called a *shears element*. It can be shown that if the spread is not desarguesian, then there is a unique shears element.

Note that this is not the only possible characterisation of a spread arising from a semifield. Another is via *reguli*; a spread is a semifield spread if and only if there exists a spread element such that every regulus defined by it and two other spread elements is contained in the spread. If every regulus defined by any three elements of the spread is contained in the spread, then the spread is desarguesian.

2.5.2 Translation Planes

Definition 2.60. A *translation* of $V(n,\mathbb{F})$ is a map $\tau_u : v \mapsto u + v$. The *translation group* is the set of all translations with composition.

Exercise 2.61. Suppose $\mathcal{S}$ is a spread, and $T(\mathcal{S})$ the corresponding affine plane. Show that

- Translations map lines to lines.
- Translations preserve parallel classes.
- Translations act transitively on each parallel class except one.

Hence these translations can be extended to collineations of the projective plane, fixing the line at infinity pointwise, and fixing all lines containing one particular spread element. These are *elations* of the plane with axis the line at infinity. A plane with a group of elations with fixed axis acting transitively on the remaining points of the plane is called a *translation plane*. This characterises the projective planes obtained from spreads, i.e. from quasifields.

Lemma 2.62. *The dual of a projective plane obtained from a quasifield isomorphic to the projective plane obtained from the opposite quasifield.*

Since a semifield is precisely a quasifield whose opposite is also a quasifield, we get the following characterisation of projective planes arising from semifields.

Definition 2.63. A *semifield plane* is a translation plane whose dual is also a translation plane.

These results are due to André [3], and Bruck and Bose [9]. This setup is usually called the ABB construction of a plane. An equivalent, more projective geometric setup, is to instead view $V(2n, q)$ as a co-dimension one subspace of $V(2n+1, q)$; or, $\mathrm{PG}(2n-1, q)$ as a hyperplane at infinity in $\mathrm{PG}(2n, q)$. We then take points to be the affine points in $\mathrm{PG}(2n, q)$, together with the spread elements, and lines to be subspaces of vector space dimension $n+1$ (projective dimension n) meeting the hyperplane at infinity in a spread element.

Chapter 3

Constructions, Classifications, and Invariants

It is not straightforward to decide on the best setting in order to construct, classify, or analyse semifields. Each of the following settings comes with its own strengths and weaknesses. For each, we should ask ourselves:

- does this setting allow us to construct some semifields easily?
- does this setting allow us to determine whether two given semifields are equivalent?
- does this setting suggest any subclasses of semifields which may be classifiable?
- does this setting suggest any useful isotopy invariants?

3.1 Representations and Constructions

3.1.1 Linearised Polynomials

Consider the finite field $\mathbb{F}_{q^n}$, and regard it as a vector space over $\mathbb{F}_q$. We can make this correspondence explicit by choosing an $\mathbb{F}_q$-basis for $\mathbb{F}_{q^n}$. However, we will try to avoid doing this as much as possible.

Consider now $\mathbb{F}_q$-linear maps from $\mathbb{F}_{q^n}$ to itself; that is, the algebra $\mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_{q^n})$ with addition and composition of maps. Again by choosing a basis, we can identify this algebra with the matrix algebra $M_n(\mathbb{F}_q)$.

Every map from a finite field to itself can be expressed as a polynomial, by interpolating. Since every element of $\mathbb{F}_{q^n}$ is a root of the polynomial $x^{q^n} - x$, each map can be expressed uniquely as a polynomial in $\mathbb{F}_{q^n}[x]$ of degree at most $q^n - 1$. It is natural to ask if we can identify which polynomials represent the $\mathbb{F}_q$-linear maps.

Definition 3.1. A *linearised polynomial* is a polynomial of the form $f(x) = \sum_{i=0}^{k} f_i x^{q^i} \in \mathbb{F}_{q^n}[x]$.

Lemma 3.2. *The algebra* $\mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_{q^n})$ *is isomorphic to the ring of linearised polynomials with addition and composition modulo* $x^{q^n} - x$.

One useful method to determine the rank of a linearised polynomial is via the *Dickson matrix*.

$$D_f := \begin{pmatrix} f_0 & f_1 & \cdots & f_{n-1} \\ f_{n-1}^q & f_0^q & \cdots & f_{n-2}^q \\ \vdots & \ddots & \ddots & \vdots \\ f_1^{q^{n-1}} & f_2^{q^{n-1}} & \cdots & f_0^{q^{n-1}} \end{pmatrix}$$

Lemma 3.3. *The rank of a linearised polynomial* $f(x)$ *as an* $\mathbb{F}_q$*-linear map on* $\mathbb{F}_{q^n}$ *is equal to the rank of the Dickson matrix* D_f.

This can (and has) been used to good effect; however it still gives a quite complicated set of conditions. In some cases we can find conditions that don't look too bad.

Exercise 3.4. Find the possible ranks of a linearised polynomial of the form $f(x) = f_0 x + f_1^{q^s}$, and conditions on the coefficients to determine the rank.

Exercise 3.5. Let $n = 3$. Find conditions on the coefficients of $f(x) = f_0 x + f_1 x^q + f_2 x^{q^2}$ which determine when $f(x)$ defines an invertible linear map on $\mathbb{F}_{q^3}$.

Exercise 3.6. Suppose $f(x)$ defines an $\mathbb{F}_{q^s}$-linear map on $\mathbb{F}_{q^n} = \mathbb{F}_{q^{ms}}$, in which case $f(x) = \sum_{i=0}^{m-1} f_{is} x^{q^{is}}$. Let $D_n 1$ denote its Dickson matrix when regarded as an $\mathbb{F}_q$-linear map, and D_2 its Dickson matrix when regarded as an $\mathbb{F}_{q^s}$-linear map. Show that $\det(D_1) = N_{\mathbb{F}_{q^s}:\mathbb{F}_q}(D_2)$.

We know from linear algebra that transposing a matrix does not affect its rank. The transpose of a Dickson matrix is again a Dickson matrix; we could ask how the associated polynomial relates to the original polynomial. In fact it turns out to be the *adjoint*.

Definition 3.7. The *adjoint* of a linearised polynomial $f(x)$ is the polynomial $\hat{f}(x) := f_0 x + \sum_{i=1}^{n-1} f_{n-i}^{q^i}$.

Exercise 3.8. Show that $\mathrm{tr}(x f(y)) = \mathrm{tr}(\hat{f}(x) y)$ for all $x, y \in \mathbb{F}_{q^n}$.

3.1.2 Bilinear Maps as Polynomials

We have seen already that linear maps can be represented as linearized polynomials. In a similar manner, we can represent bilinear maps (i.e. multiplications) as polynomials in two variables of a special shape.

Lemma 3.9. *Let $\circ$ be a bilinear map from $\mathbb{F}_{q^n}$ to itself. Then there exist unique $c_{ij} \in \mathbb{F}_{q^n}$ such that*

$$x \circ y = \sum_{i,j=0}^{n-1} c_{ij} x^{q^i} y^{q^j}.$$

Thus we can represent a multiplication by an $n \times n$ matrix with entries in $\mathbb{F}_{q^n}$.

We can gather terms together to write another (sometimes) convenient representation.

Lemma 3.10. *Let $\circ$ be a bilinear map from $\mathbb{F}_{q^n}$ to itself. Then there exist unique linearised polynmials $c_i(y)$ such that*

$$x \circ y = \sum_{i=0}^{n-1} c_i(y) x^{q^i}.$$

Though these representations are neat, it is usually quite difficult to determine whether or not a multiplication represented in this way defines a presemifield. Usually we need to restrict the shape further, say by assuming not too many of the c_{ij} or $c_i(y)$ are not zero. For the former, the Generalised Twisted Fields are a good example.

Exercise 3.11. Show that the following defines a presemifield.

$$x \circ y = xy + (\mathrm{tr}(x)y + \mathrm{tr}(y)x)^2.$$

These are known as *Knuth's binary semifields*.

The Knuth orbit can be seen in this setting by considering the trilinear form

$$f(x, y, z) = \mathrm{tr}((x \circ y)z).$$

3.1.3 Semifields with a large nucleus

One of the most studied cases of semifields are those which are two-dimensional over a nucleus, sometimes referred to as *rank two* semifields. This is well-studied for a variety of reasons, with one of the main reasons being that it is simply easier to construct and analyse these semifields, due to the large associative component.

Suppose $n = 2m$, and $\mathbb{S}$ is a semifield of order q^{2m} with centre containing $\mathbb{F}_q$ and a nucleus of order q^m. Up to Knuth equivalence, we can assume without loss of generality that this is the right nucleus.

From Section 2.3.1, we hence have that the elements of $C(\mathbb{S})$ commute with every element of a subfield of $M_{2m}(\mathbb{F}_q)$ isomorphic to $\mathbb{F}_{q^m}$. Let us fix this subfield to be

$$C(\mathbb{F}_{q^m}) := \{\alpha x : \alpha \in \mathbb{F}_{q^m}\}$$

regarded as a subset of linearized polynomials over $\mathbb{F}_{q^{2m}}$.

We have that $C(\mathbb{F}_{q^m}) \subseteq \mathrm{Cent}(C(\mathbb{S}))$, and hence $C(\mathbb{S})) \subseteq \mathrm{Cent}(C(\mathbb{F}_{q^m}))$.

Exercise 3.12. Suppose $f(x) \in \mathrm{Cent}(C(\mathbb{F}_{q^m}))$. Show that $f(x) = f_0 x + f_m x^{q^m}$ for some $f_0, f_m \in \mathbb{F}_{q^{2m}}$.

Exercise 3.13. Let $f(x) = f_0 x + f_m x^{q^m}$ for some $f_0, f_m \in \mathbb{F}_{q^{2m}}$. Show that $f(x)$ is invertible if and only if $f_0^{q^m+1} - f_m^{q^m+1} \neq 0$.

Exercise 3.14. Show that there exist $\mathbb{F}_q$-linearised polynomials maps $a(y), b(y)$ over $\mathbb{F}_{q^{2m}}$ such that

$$x \circ y = a(y)x + b(y)x^{q^m}$$

such that

$$a(y)^{q^m+1} - b(y)^{q^m+1} \neq 0$$

for all $y \neq 0$.

Exercise 3.15. Let $\{1, \theta\}$ be an $\mathbb{F}_{q^m}$-basis for $\mathbb{F}_{q^{2m}}$. Write any $y \in \mathbb{F} + q^{2m}$ as $y = y_0 + y_1\theta$, and let $a(y) = y_0$, $b(y) = \epsilon y_1$. Determine conditions for which the multiplication

$$x \circ y = a(y)x + b(y)x^{q^m}$$

defines a presemifield. Can you write a and b as linearised polynomials?

These are the *Hughes-Kleinfeld semifields* [24].

3.1.4 BEL rank

Both the generalised twisted fields and the semifields of rank two over a nucleus can be written in the following way:

$$x \circ y = f_1(x)g_1(y) + f_2(x)g_2(y),$$

where f_i, g_i are linearised polynomials. Can any other semifields be expressed in this way?

It turns out that semifields with such a multplication are precisely those which can be constructed from a *BEL configuration* in $V(2n, q)$. This is a geometric construction from Ball, Ebert, and Lavrauw [5] (2007) which uses two subspaces of $V(kn, q)$ of complementary dimension such that no element of a fixed desarguesian spread meets both nontrivially. In [36], Michel Lavrauw and I defined the *BEL rank* of a semifield as the minimal k for which there exist linearised polynomials f_i, g_i such that $\mathbb{S}$ is isotopic to a semifield with multiplication

$$x \circ y = \sum_{i=1}^{k} f_k(x)g_(y).$$

Taking $(f_1(x), f_2(x)) = (x, x^{q^{n/2}})$ gives precisely the case of semifields rank two over a nucleus.

What if we take $(f_1(x), f_2(x)) = (x, x^{q^s})$? If $\gcd(s, n) = 1$, then results of Carlitz and McDonnell [42] can be used to show that the semifield is a generalised twisted field.

However in [36] it was shown that there exists a semifield of order 2^6 with multiplication

$$x \circ y = xg_1(y) + x^{q^2} g_2(y),$$

not isotopic to a twisted field; indeed, not isotopic to any previously known construction.

What if we take $(f_1(x), f_2(x)) = (x, \mathrm{tr}_{q^n:q^s}(x))$? Again in [36] it was shown that there exists a semifield of order 2^6 with multiplication

$$x \circ y = xg_1(y) + (x + x^{q^2} + x^{q^4})g_2(y),$$

not isotopic to a twisted field; indeed, not isotopic to any previously known construction.

Open Problem 3.1.1. Can these examples be generalised to other q, or other n?

3.2 Biprojective Polynomials

Let us identify the elements of $\mathbb{S}$ with tuples over a field; say, elements of $(\mathbb{F}_{q^m})^s$, and suppose the centre of the field contains $\mathbb{F}_q$.

$$x \circ y = (x_1, \ldots, x_s) \circ (y_1, \ldots, y_s)$$

In particular let us take $s = 2$. Then we have

$$(x_1, x_2) \circ (y_1, y_2) = (f_1(x_1, x_2, y_1, y_2), f_2(x_1, x_2, y_1, y_2))$$

for some maps f_1, f_2.

For example, the semifields of Dickson from Exercise 2.12 is usually expressed in this way.

Very recently, Lukas Kölsch and Faruk Göloğlu [21] have shown that many of these constructions can be written in a particularly nice way using *biprojective polynomials*, and used this representation to construct a very large family of new semifields.

$$\begin{aligned}(x, y) * (y, v) = ((a_0u + b_0v)x^q + (a_0u^q + c_0v^q)x + (c_0u + d_0v)y^q + (b_0u^q + d_0v^q)y,\\ (a_1u + b_1v)x^r + (a_1u^r + c_1v^r)x + (c_1u + d_1v)y^r + (b_1u^r + d_1v^r)y).\end{aligned}$$

Semifields of this form have a very nice property; their autotopism groups contain a large cyclic group, namely one of order $q^m - 1$. Notably, this is the same

as the maximum possible size of a nucleus; however, the semifields constructed can have trivial nuclei while still having a large cyclic subgroup of the autotopism group. Moreover, the full autotopism group appears to be contained in $\Gamma\mathrm{L}(2, q^m)^3$; we saw that the generalised twisted fields have their autotopism group contained in $\Gamma\mathrm{L}(1, q^{2m})^3$, and so this is in some sense the next simplest case.

3.3 Cyclic Semifields and Skew Polynomial Rings

The classical construction of a finite field involves taking a polynomial ring modulo an irreducible. In order to generalise this, one could seek to mimic this construction using a more general type of polynomial ring. One such ring is a *skew polynomial ring*.

Definition 3.16. Let σ be an $\mathbb{F}_q$-automorphism of $\mathbb{F}_{q^m}$. The *skew polynomial ring* $\mathbb{F}_{q^m}[t;\sigma]$ is defined as the set of polynomials in t with coefficients in $\mathbb{F}_{q^m}$, where addition is usual polynomial addition, and multiplication is satisfies the usual rules except for the field elements commuting with the indeterminate; instead we have that

$$t\alpha = \alpha^\sigma t$$

for all $\alpha \in \mathbb{F}_{q^m}$.

Remark 3.17. This structure is also referred to in the literature as a *twisted polynomial ring*, or a *twisted group algebra*. The general definition of a skew-polynomial ring can also involve a *derivation*; however over finite fields we can assume up to isomorphism that this is zero.

When $x^\sigma = x^q$, we can identify the skew-polynomial ring with the ring of linearised polynomials with composition; we identify x^{q^i} with t^i. The isomorphism follows from the fact that $x^q \circ (\alpha x) = \alpha^q x^q = (\alpha^q x) \circ x^q$. Note here that $\circ$ denotes the composition of maps, rather than a semifield multiplication!

Exercise 3.18. Verify the following in $\mathbb{F}_4[t,\sigma]$. Here $\alpha^2 + \alpha + 1 = 0$.

Let $f = t^2 + \alpha t + 1$, $g = t + \alpha$. Then:

- $fg = t^3 + \alpha$,
- $gf = t^3 + t^2 + \alpha t + \alpha$
- $f = (t+1)g + \alpha^2 = gt + 1$
- $t^2 + 1 = (t+1)(t+1) = (t+\alpha)(t+\alpha^2) = (t+\alpha^2)(t+\alpha)$

It turns out that skew-polynomial rings possess many of the key properties of polynomial rings. Although multiplication is not commutative, we do have a

(left- and right-)Euclidean algorithm which allows us to define remainders. Furthermore while factorisation into irreducibles is not unique, the multiset of the degrees in a factorisation into irreducibles is unique. These properties allow us to define a semifield multiplication.

Lemma 3.19. *Let $f(t)$ be irreducible in $\mathbb{F}_{q^m}[t;\sigma]$ of degree s, and let $\mathbb{S}$ be the algebra with elements in $\mathbb{F}_{q^m}[t;\sigma]$ of degree less than s and multiplication defined by*

$$a(t) \circ b(t) := a(t)b(t) \mod_r f(t).$$

Then $\mathbb{S}$ is a semifield.

This was first discovered by Petit in 1965 [46]; however it was not well-known in the semifield community. An equivalent construction using semilinear maps was discovered by Jha and Johnson (1989) [25]; these became known as *cyclic semifields*, since they can be seen as a generalisation of *cyclic algebras*.

Exercise 3.20. Find an irreducible skew-polynomial of degree 2 in $\mathbb{F}_4[t,\sigma]$.

Exercise 3.21. Suppose $f(t) = t^2 + \lambda t + \mu$ is irreducible in $\mathbb{F}_{q^2}[t,\sigma]$. Write the multiplication in the semifield defined above in terms of the coefficients of $a(t)$ and $b(t)$.

The question of isotopy classification and autotopy groups of these semifields has been studied by [30], [19], [35]. More recently, I was able to extend this construction to a larger family of semifields (and MRD codes) [51]. The constructions used a couple of key lemmas; for us, the most useful are the following.

Lemma 3.22. *Suppose $F(y) \in \mathbb{F}_q[y]$ is irreducible of degree s. Then any factor of $F(t^m)$ in $\mathbb{F}_{q^m}[t;\sigma]$ has degree divisible by s.*

Lemma 3.23. *Suppose $F(y) \in \mathbb{F}_q[y]$ is irreducible of degree s. Then $(F(t^m))$ is a maximal two-sided ideal in $\mathbb{F}_{q^m}[t;\sigma]$, and*

$$\frac{\mathbb{F}_{q^m}[t;\sigma]}{(F(t^m))} \simeq M_m(\mathbb{F}_{q^s}),$$

where $\simeq$ denotes an isomorphism as algebras. Moreover, the rank of the image of a skew-polynomial $f(t)$ satisfies

$$\operatorname{rank}(f(t)) = m - \frac{\deg(\operatorname{gcrd}(f(t), F(t^m)))}{s}.$$

From this, we get that the image of the set of polynomials of degree less than sk is an additively closed set of matrices in which every nonzero element has degree at most $m-k+1$. Taking $k=1$ returns that this is a semifield spread set. Moreover, since the set of polynomials is closed under $\mathbb{F}_{q^m}$-multplication, the spread set has left-idealiser containing $\mathbb{F}_{q^m}$, and so the left nucleus of the semifield contains $\mathbb{F}_{q^m}$.

Remark 3.24. In [51], a more general construction containing semifields with a broader range of parameters was constructed. However the construction is slightly beyond the scope of these notes. We will just note that the construction is a hybrid of the construction for cyclic semifields and for generalised twisted fields, and contains both as special cases.

3.3.1 Other Connections

For instance, flocks.

3.3.2 Kantor's Construction from Spreads

While spreads are interesting objects, usually it is difficult to construct them directly. However in [29], a construction for commutative semifields of even order was found using spreads directly. Moreover, the number of semifields produced grows more quickly than any other known family; it is not bounded above by any polynomial.

3.3.3 Scattered Subspaces and Covering Radius

At first look, spreads arising from semifields look more or less the same. They (by definition) have the same intersection properties, and possess many of the the same automorphisms. Apart from the full automorphism group, it seems difficult to find methods to distinguis spreads apart directly.

However, we can define spread-based invariants by considering how other subspaces can intersect a given spread.

Definition 3.25. A subspace U is said to be $(\mathcal{S}, h)$-*scattered* if

$$\dim(U \cap S) \leq h \quad \text{for all } S \in \mathcal{S}.$$

The concept of scattered subspaces was first introduced in [4], [7]. More recently generalisations were considered in for example [6], [23].

We can then consider the maximum dimension of a $(\mathcal{S}, h)$-scattered subspace, or the minimum h for which a $(\mathcal{S}, h)$-scattered subspace of a fixed dimension exists. In [2], [23] this was related to the notion of the *covering radius* of the semifield spread set and its inverse.

In the case $q = 2$. the existence of a $(\mathcal{S}, 1)$-scattered subspace of dimension n corresponds to the existence of a translation hyperoval in the corresponding translation plane.

In [2], Kevin Allen and I demonstrated the first example of a semifield (or indeed translation) plane without a translation hyperoval. Equivalently, we demon-

strated that semifield spread can possess different properties with respect to scatteredness. This is a first step; there are many more questions left to explore.

3.4 Classifications

There are many constructionso of semifields; in general, it is thought that full classification is impossible. However with certain restrictions, some classifications are possible. We start with some classifications for small dimension over the centre.

Exercise 3.26. Show that every semifield two-dimensional over its centre is isotopic to a field. **Dickson's Theorem**.

Menichetti has proved the strongest classificaiton results to date. He first showed the following in the [43] in 1977.

Theorem 3.27. *Let $\mathbb{S}$ be a semifield three-dimensional over its centre. Then $\mathbb{S}$ is isotopic to a field or a generalised twisted field.*

Some years later [44] (1996) he extended this further.

Theorem 3.28. *Let $\mathbb{S}$ be a semifield n-dimensional over its centre $\mathbb{F}_q$, for n prime, and suppose that q is large enough with respect to n. Then $\mathbb{S}$ is isotopic to a field or a generalised twisted field.*

The method used by Menichetti in this second paper was to study a *determinantal hypersurface* related to $\mathbb{S}$.

Definition 3.29. Let $\mathbb{S}$ be a semifield of dimension n over $\mathbb{F}_q$ with $\mathbb{F}_q$ contained in its centre, and let $\{E_1, \ldots, E_n\}$ be an $\mathbb{F}_q$-basis for $C(\mathbb{S})$. Then define the *determinantal polynomial* $f_{\mathbb{S}}$ of $\mathbb{S}$ as

$$f_{\mathbb{S}}(x_1, \ldots, x_n) := \det\left(\sum_i x_i E_i\right).$$

This is a polynomial in n variables over $\mathbb{F}_q$ and has degree n.

Lemma 3.30. *The polynomial $f_{\mathbb{S}}$ is irreducible over $\mathbb{F}_q$. If q is large enough with respect to n, then it is reducible over some field $\mathbb{F}_{q^s}$ for some s (not absolutely irreducible). Moreover, s divides n, and $f_{\mathbb{S}} = gg^{\sigma} \cdots g^{\sigma^{s-1}}$ for some polynomial g over $\mathbb{F}_{q^s}$.*

When n is prime, we must have $s = n$, and so $f_{\mathbb{S}}$ is a product of linear polynomials over $\mathbb{F}_{q^n}$. What Menichetti did was to show that any semifield whose determinantal polynomial factorises in this way must be isotopic to a field or twisted field.

For n not prime, one must consider the possiblity that $f_{\mathbb{S}}$ factorises over an intermediate field; for example, with $n = 4$ we could have $s = 2$, and so $f_{\mathbb{S}}$ could be a product of two (conjugate) quadratic polynomials in four variables over $\mathbb{F}_{q^2}$. Some results have been obtained (e.g. Bani-Ata et al, Rua), but nothing as strong as Menichetti's results has been found yet.

Exercise 3.31. Show that if $\mathbb{S}$ has nucleus containing $\mathbb{F}_{q^s}$, then $f_{\mathbb{S}}$ factorises over $\mathbb{F}_{q^s}$.

Note that the converse is not true in general; however, we don't know exactly when this can occur (beyond the case of twisted fields).

3.4.1 Computer Classifications

Over the years as computational power grew, full classifications of semifields of certain orders have been performed. However even with computations distributed across large clusters, we have only reached a handful of cases.

n	q	#Classes	Reference
4	2	3 (3)	Knuth 1965 [31]
4	3	27 (12)	Dempwolff 2008 [18]
4	4	(28)	Rua et al 2011[14]
4	5	(42)	Rua et al 2011 [14]
4	7	(120)	Rua et al 2012[15]
5	2	6 (3)	Walker 1962 [52]
5	3	23 (9)	Rua et al 2012 [50]
6	2	332 (80)	Rua et al 2009 [49]

x=number of isotopy classes, (x)=number of Knuth orbits

Of the 332 isotopy classes of order 2^6, only 35 were from known constructions.

Many of the known constructions work only in composite dimension. For prime dimension, we know that for q large enough the number of semifields is low. It would be very interesting to determine the number of isotopy classes of semifields of order 2^7, to determine whether or not this behaviour only kicks in for large q.

A classification for 3^6 would also be interesting, to see whether the large number of semifields of order 2^6 is an anomaly, or the typical behaviour.

3.5 Other Topics

3.5.1 Fractional Semifields

There are many other curious properties that semifields can have, and many questions about them that one can explore.

For example, it is well known that a field $\mathbb{F}_{q^m}$ is contained in $\mathbb{F}_{q^n}$ if and only if m divides n. This is because a field is always a vector space over any subfield. However, for semifields this is not the case. We can find, for example, semifields of order q^5 containing semifields of order q^2. The larger semifield has *fractional dimension* over the smaller semifield. These have been explored for example in [26].

3.5.2 Non-primitive semifields

We all know that the multplicative group of a finite field is cyclic; it contains *primitive elements*. For semifields, the multiplicative structure is not a group, but rather a *loop*. We must take care if we want a sensible definition of a primitive element, since powers are not well-defined. However if we define powers recursively by always multiplying on the left (or right, so long as we stick to the same side every time), then the definition does make sense. Rua et al [47] showed that not every semifield contains a primitive element. Rod Gow and I showed that every twisted field contains a primitive element [22]. Rua has some further results for four-dimensional semifields [48]. Beyond that, the field is wide open.

Chapter 4

The Geometry of Matrices and Tensors

4.1 Matrices and Linear Maps

Matrices, as it is well known, represent linear maps between finite vector spaces, or bilinear maps from a direct product of vector spaces to the underlying field, i.e. bilinear forms.

$$M_{m\times n}(F) \leftrightarrow \mathrm{Hom}_F(V,W) \leftrightarrow \mathrm{Bil}(V\times W, F)$$

The space of linear maps is partitioned into sets according to their *rank*, the usual linear algebraic definition being

$$\mathrm{rank}(A) = \dim(\mathrm{Im}(A)).$$

Lemma 4.1. *The set of matrices of rank one is equal to the set* $S_{m,n} := \{vw^T : v \in V = F^m, w \in W = F^n\}$.

Moreover, the rank of a nonzero matrix A *is equal to the minimum positive integer* r *such that* A *can be written as the sum of* r *matrices of rank one.*

The maximum rank of a matrix is $\max\{m,n\}$, and there exists a matrix of every possible rank between 1 and this maximum.

When $m = m$, a matrix is invertible if and only if its determinant is non-zero, if and only if it has rank n.

The rank of a matrix is equal to the size of the smallest non-zero minor.

We can also view matrices as *tensors*; that is, the space $V \otimes W$. This space is defined as a vector space spanned by symbols of the form $v \otimes w$ for $v \in V, w \in W$, subject to the rules $(v + \lambda v') \otimes w = v \otimes w + \lambda(v' \otimes w)$ for $v, v' \in V$, $w \in W$, $\lambda \in F$. We can make this correspondence explicit by identifying vw^T with $v \otimes w$.

Since this product behaves nicely with respect to scalar multiplication, we can also view this as an embedding of $\mathrm{PG}(m-1,F)\times\mathrm{PG}(n-1,F)$ into $\mathrm{PG}(mn-1,F)$. This is the *Segre embedding*, and its image (corresponding to points defined by the rank one matrices) is the *Segre variety*.

Note that this is named after Corrado Segre; not the Beniamino Segre who is perhaps more well-known due to his work on conics (and much more).

The use of the word *variety* can be justified by observing that the matrices of rank one is equal to the set of matrices for which every 2×2 minor is zero; each 2×2 minor is a quadratic form, and so the matrices of rank one are the intersection of a bunch of (very degenerate, if $(m,n) \neq (2,2)$) quadrics.

In this setting, the rank of a matrix corresponds to the smallest secant variety with respect to the Segre variety in which the point defined by the matrix lies.

It is well known that the additive rank-preserving maps are those of the form

$$A \mapsto XA^{\rho}Y$$

for X, Y invertible and ρ a field automorphism, unless $m = n$ (in which case we just need to include transposition). It is clear that each such map preserves the Segre variety, and hence preserves rank. The converse takes more thought; it is usually attributed to Hua and Wan [53]. Various other questions regarding additive maps preserving certain sets with specified rank properties have been studied over the years; some interesting questions remain.

Furthermore it is well-known that this group, the stabiliser of the Segre variety, acts transitively on matrices of the same rank. Hence there are precisely n orbits of points of $\mathrm{PG}(mn-1,F)$ under this group.

4.1.1 Delsarte Duality

Definition 4.2. Let C be an $\mathbb{F}_q$-subspace of $M_m(\mathbb{F}_{q^s})$. The *Delsarte dual* of C is defined as

$$C^{\perp} := \{Y \in M_m(\mathbb{F}_{q^s}) | \, \mathrm{tr}(\mathrm{Tr}(XY^T)) = 0 \, \forall \, X \in C\}.$$

Here Tr denotes matrix trace, while tr denotes field trace from $\mathbb{F}_{q^s}$ to $\mathbb{F}_q$.

Delsarte showed the following in [17].

Theorem 4.3. *Suppose C is an $\mathbb{F}_q$-subspace of $M_m(\mathbb{F}_{q^s})$ such that every nonzero element of C has rank at least k. Then*

$$\dim_{\mathbb{F}_q}(C) \leq sm(m-k+1).$$

Moreover, in the case of equality, then every nonzero element of $\mathbb{C}^{\perp}$ has rank at least $n-k+2$.

In the languate of rank metric codes, this states that the Delsarte dual of an MRD code is again an MRD code.

For semifields, this tells us that C is a semifield spread set if and only if its dual contains no elements of rank one.

4.1.2 Tensors

We can define the tensor product of more than two vector spaces, either recursively or by a direct definition.

In this case, we can again define a Segre variety and the rank of a tensor; in this case the rank is defined using the characterisation from Lemma 4.1.

We can view tensors as multidimensional arrays of field elements, in the same way as we do for matrices. Tensors define multilinear maps and multilinear forms in an analogous way. In particular, 3-fold tensors correspond to bilinear maps, i.e. multiplications of (not necessarily associative) algebras!

Despite tensors being analogous to matrices and having been studied for almost as long, the understanding of the ranks, orbits, and geometry with respect to the Segre variety is far from known. For example, it is not known in general what the maximum rank of a tensor in a given space is. It is not known whether or not *nonsingular tensors*, the analogue of invertible matrices, must have rank greater than or equal to any singular tensor. It is known that there exist nonsingular tensors with the same rank as a singular tensor in the same space, and it is also known that two nonsingular tensors can have different rank.

Consider a 3-fold tensor in $V \otimes V \otimes V$, where $V = V(n, q)$. Given an element $a^\vee$ of the dual space $V^\vee$, and a pure tensor $T = v_1 \otimes v_2 \otimes v_3$, we can define

$$a^\vee(T) = T(a^\vee) = a^\vee(v_1)v_2 \otimes v_3 \in V \otimes V.$$

Thus the contraction of a 3-tensor is a 2-tensor, which can be represented as a matrix.

We can also extend this linearly to any sum of pure tensors.

Definition 4.4. The set $C(T) := \{a^\vee(T) : a^\vee \in V^\vee\}$ is called the (first) *contraction space* of T. It is an $\mathbb{F}_q$-subspace of $V \otimes V \simeq M_n(\mathbb{F}_q)$.

Definition 4.5. A tensor is *nonsingular* if every nontrivial contraction of it is nonsingular.

Hence a tensor is nonsingular if and only if every nonzero element of $C(T)$ is a nonsingular matrix, and $C(T)$ has dimension n; in other words, if and only if $C(T)$ is a the spread set of a semifield with centre containing $\mathbb{F}_q$.

4.2 Linear Sets and Semifields

Suppose $\mathbb{S}$ is a semifield of order q^{ms} with centre $\mathbb{F}_q$ and left nucleus containing $\mathbb{F}_{q^s}$. Then its spread set C is an $\mathbb{F}_q$-subspace of $M_m(\mathbb{F}_{q^s})$. Two semifields with these properties are equivalent if and only if they are equivalent via

$$C' = XC^\rho Y,$$

where $X, Y \in \mathrm{GL}(m, q^s)$ and $\rho \in \mathrm{Aut}(\mathbb{F}_{q^s})$.

Since any such equivalence preserves $\mathbb{F}_{q^s}$-subspaces, the geometry of intersections between C and $\mathbb{F}_{q^s}$-subspaces is an isotopy invariant.

Definition 4.6. Let $\mathrm{Gr}(n, k, q)$ denote the set of k-dimensional subspaces of an n-dimensional vector space over $\mathbb{F}_q$. Or, equivalently, the set of $(k-1)$-dimensional subspaces of an $(n-1)$-dimensional projective space $\mathrm{PG}(n-1, q)$.

$\mathrm{Gr}(n, 1, q)$	points of $\mathrm{PG}(n-1, q)$
$\mathrm{Gr}(n, 2, q)$	lines of $\mathrm{PG}(n-1, q)$
$\mathrm{Gr}(n, 3, q)$	planes of $\mathrm{PG}(n-1, q)$
$\mathrm{Gr}(n, k, q)$	$(k-1)$-spaces of $\mathrm{PG}(n-1, q)$

Projectively speaking, we consider $L(C)$ a *linear set* in $\mathrm{PG}(m^2-1, q^s)$. If $\mathbb{S}$ and $\mathbb{S}'$ are isotopic, then $L(C)$ and $L(C')$ are equivalent under a collineation of $\mathrm{PG}(m^2-1, q^s)$. Note however that the converse is not necessarily true. Although papers addressing this topic usually use the language of projective geometry and linear sets, the actual calculations usually require us to work in the underlying vector space, so we will stick to vector space notation for now (borrowing terminology from projective geometry when convenient).

Definition 4.7. The *weight* of an $\mathbb{F}_{q^s}$-subspace W of $V(n, q^s)$ with respect to an $\mathbb{F}_q$-subspace U of $V(n, q^s)$ is defined as

$$w_U(W) := \dim(U \cap W)_{\mathbb{F}_q}.$$

Definition 4.8. The *k-th weight distribution* of an $\mathbb{F}_q$-subspace U of $V(n, q^s)$ is the multiset

$$w_U(k) := \{^*\dim(U \cap W)_{\mathbb{F}_q} : W \in \mathrm{Gr}(n, k, q^s)^*\}.$$

Since every rank-preserving map on $M_m(\mathbb{F}_{q^s})$ preserves each $\mathrm{Gr}(n, k, q^s)$, the weight distributions of a spread set of a semifield with a nucleus containing $\mathbb{F}_{q^s}$ and centre containing $\mathbb{F}_q$ is invariant under isotopy.

Lemma 4.9. *Suppose $C = C(\mathbb{S})$ and $C' = C(\mathbb{S}')$ are both $\mathbb{F}_q$-subspaces of $M_m(\mathbb{F}_{q^s})$. Then $w_C(k) = w_{C'}(k)$ for all k.*

This can sometimes allow us to show that some semifields with the same nuclei are inequivalent.

4.2.1 The Geometry of Quadrics in $\mathrm{PG}(3, q^s)$

Thought the weight distribution can be useful, we are ignoring much of the geometry at our disposal. In particular, we are not exploiting the fact that rank-preserving maps not only fix each $\mathrm{Gr}(n, k, q^s)$, but also fix the Segre variety (and each of its secant varieties).

Consider the case $m = 2$; that is, semifields which are two-dimensional over a nucleus. We are thus considering $\mathbb{F}_q$-subspaces of $M_2(\mathbb{F}_{q^s})$ of dimension $2s$ disjoint from the set of rank one matrices. Projectively, this corresponds to *linear sets* in $\mathrm{PG}(3, q^s)$ defined by $\mathbb{F}_q$-subspaces of dimension $2s$, disjoint from a hyperbolic quadric $Q^+(3, q^s)$ determined by the equation $\det(X) = x_{11}x_{22} - x_{12}x_{21} = 0$. The equivalence becomes equivalence under the stabiliser of the quadric.

In order to construct or classify such semifields, we have two options:

- Fix a quadric and construct/classify linear sets disjoint from it;
- Classify linear sets, and attempt to find and classify hyperbolic quadrics disjoint from it.

It is well-known that every nondegenerate quadric determines a *polarity* of the projective space; that is, a map $\perp$ of order two sending $\mathrm{Gr}(n, k, q^s)$ to $\mathrm{Gr}(n, n-k, q^s)$ for each k which reverses inclusion (i.e. $U \leq W \Leftrightarrow W^\perp \leq U^\perp$. This can be realised by using the symmetric bilinear form $b(u, w) := Q(u+w) - Q(u) - Q(w)$. Then $U^\perp = \{w : b(u, w) = 0 \; \forall \; u \in U\}$.

We can extend this to a map on $\mathbb{F}_q$-subspaces by defining

$$U^\perp = \{w : \mathrm{tr}(b(u, w)) = 0 \; \forall \; u \in U\},$$

where tr denotes the field trace from $\mathbb{F}_{q^s}$ to $\mathbb{F}_q$. Note that if U is an $\mathbb{F}_{q^s}$-subspace, then this coincides with the previous definition of $\perp$.

Suppose U is an $\mathbb{F}_q$-subspace of $M_2(\mathbb{F}_{q^s})$ of $\mathbb{F}_q$-dimension $2s$ disjoint from the set of rank-one matrices; projectively speaking, disjoint from the $Q^+(3, q^s)$ defined above. What can we say about $U^\perp$? It is again an $\mathbb{F}_q$-subspace of $M_2(\mathbb{F}_{q^s})$ of $\mathbb{F}_q$-dimension $2s$. Can it contain an element of rank one?

Suppose X is a matrix of rank one in $U^\perp$ (and so $\langle X \rangle_{\mathbb{F}_{q^s}} \in Q^+(3, q^s)$). Then U is contained in $X^\perp$. But it can be checked that $X^\perp$ contains two 2-dimensional $\mathbb{F}_{q^s}$-subspaces ℓ_1, ℓ_2 consisting of rank one matrices (and zero); geometrically, it contains the two lines of $Q^+(3, q^s)$ containing $\langle X \rangle_{\mathbb{F}_{q^s}}$. But since U has $\mathbb{F}_q$-dimension $2s$, and ℓ_i has $\mathbb{F}_q$-dimension $2s$, and both are contained in the $(4s-1)$-dimensional space $X^\perp$, U and ℓ_i must meet nontrivially. But this contradicts the fact that U contains no rank one matrices. Hence we have a geometric proof of the following.

Lemma 4.10. *Suppose $C \subset M_2(\mathbb{F}_{q^s})$ is the spread set of a semifield of order q^{2s} with centre containing $\mathbb{F}_q$. Then $C^\perp$ is also the spread set of a semifield of order q^{2s} with centre containing $\mathbb{F}_q$*

This operation was introduced in [38], where it was called the *translation dual* of a semifield. In fact, it can be seen as a special case of Delsarte duality! However it is still very interesting due to its geometric interpretation and usefulness in classifying semifelds two-dimensional over a nucleus.

4.2.2 Four-dimensional semifelds, two-dimensional over a nucleus

Let us consider the possibilities for $s = 2$; this case was fully resolved by Cardinali-Polverino-Trombetti [10] (2006). In this case we are studying semifields four-dimensional over their centre and two-dimensional over a nucleus. What are the possible intersection properties of a 4-dimensional vector subspace of $M_2(\mathbb{F}_{q^2})$ with $\mathbb{F}_{q^2}$-subspaces? Or in other words, how can a linear set of rank four in $\mathrm{PG}(3, q^2)$ behave with respect to points, lines, and planes of $\mathrm{PG}(3, q^2)$, in particular if it is disjoint from a fixed hyperbolic quadric?

The possibilities are:

1. fully contained in a line;
2. meets a line in a linear set of rank three;
3. meets a point in a linear set of rank two, but not fully contained in a line;
4. meets each point in a linear set of rank at most one.

In case (1), C is in fact an $\mathbb{F}_{q^2}$-subspace, and so $\mathbb{F}_{q^2}$ is contained in the centre of $\mathbb{S}$. By Dickson's result, $\mathbb{S}$ is in fact isotopic to a field. Note that this implies (or follows from the fact) that the stabiliser of $Q^+(3, q^2)$ acts transitively on lines external to it.

In case (2), by the previous paragraph we can assume without loss of generality that the line of weight three is any fixed line external to $Q^+(3, q^2)$; the line must be external to $Q^+(3, q^2)$, for a linear set of rank three in a line $\mathrm{PG}(1, q^2)$ meets all points of the line.

Since a line contains $q^2 + 1$ points, and a plane contains $q^2 + q + 1$ points, we must have that $L(C)$ contains at least one full point. It cannot contain two full points, for then it would be equal to case (1).

4.2.3 Six-dimensional semifelds, two-dimensional over a nucleus

Consider now the case $s = 3$. This case was studied extensively in a series of papers earlier this century. In [41], it was shown that the linear set $L(C)$ of every semifield spread set in $M_2(\mathbb{F}_{q^3})$ must have one of the following geometric properties.

(0) $L(C)$ is a union of either $q^2 + q + 1$ or $q^2 + 1$ lines of a pencil.

(1) $L(C)$ is a union of q^2+q+1 lines in a plane not belonging to a pencil.

(2) $L(C)$ is a union of q^2+q+1 lines through a point, not all lines in the same plane.

(3) $L(C)$ contains a unique point of weight 2, does not contain any line and is not contained in a plane.

(4) $L(C)$ contains exactly one line and such a line contains $q+1$ points of weight 2.

(4) Any point of $L(C)$ has weight 1.

Further geometric invariants can be found. For some of these cases one can show the existence of two distinguished lines associated to a *pseudoregulus*; the positions of thse lines with respect to $Q^+(3,q^2)$ are then isotopy invariants, and so can further subdivide some cases.

For type (4), we have a distinguished line, which we will denote by ℓ. We saw previously that $Q^+(3,q^3)$ posesses a natural polarity. This case can then be further subdivided by considering the intersection of $L(C)$ and $\ell^\perp$, as in [27].

(4a) $|r^\perp \cap L(C)| = 0$

(4b) $|r^\perp \cap L(C)| = 1$

(4c) $|r^\perp \cap L(C)| = q+1$

Amongst these classes, some have full classifications; some have constructions but no classification; and some have no constructions yet known, nor a reason why they can't exist.

Semifields of order q^6 with $|\mathbb{N}_l| = q^3$ and $|\mathbb{K}| = q$

Family	$\lvert\mathbb{N}_m\rvert$	$\lvert\mathbb{N}_r\rvert$	Existence results
$\mathcal{F}_0$	q	q	**Generalized Dickson semifields, q odd**
$\mathcal{F}_1$	q	q	**Semifields from Payne–Thas ovoid of $Q(4,3^3)$**
$\mathcal{F}_2$	q	q	**Semifields from Ganley flock of $PG(3,3^3)$**
$\mathcal{F}_3$	q	q	HJ semifields [9] of type II, III, IV, V for $q=2$ Dempwolff
$\mathcal{F}_4^{(a)}$	q^2	q	$\exists$! **semifield for q odd**, $\nexists$ **semifields for q even**
	q	q^2	$\exists$! **semifield for q odd**, $\nexists$ **semifields for q even**
	q	q	?
$\mathcal{F}_4^{(b)}$	q	q	There exist semifields for $q=3$ [7]
$\mathcal{F}_4^{(c)}$	q	q	There exist semifields for any q [6]
	q^2	q^2	**Cyclic semifields for any q**
$\mathcal{F}_5$	q^3	q^3	**Hughes Kleinfeld semifields**
	q^3	q	**Knuth semifields of type (17) for any q**
	q	q^3	**Knuth semifields of type (19) for any q**
	q^2	q	$\exists$ for any $q \neq 2$ (e.g. Generalized Twisted Fields)
	q	q^2	$\exists$ for any $q \neq 2$ (e.g. Generalized Twisted Fields)
	q	q	$\exists$ for any $q \neq 2$ (e.g. Generalized Twisted Fields)

This shows the difficulty of classifying semifields, even in small dimensions. However it also highlights how some particular classes of semifields can be characterised or classified; for example, a full geometric characterisation of cyclic semifields has been acheived in this dimension. Perhaps there is scope to explore certain cases in higher dimensions.

4.3 Rank Two and Three Commutative Semifields

Semifields which are both commutative and two-dimensional over their left nucleus have been partially classified. Using techniques from both finite geometry and algebraic geometry, it was shown in [8] and [33] that if q is large enough with respect to n, then all such semifields are isotopic to a handful of known examples. There remains space to improve this; the proof uses only a small part of the picture.

Semifields which are commutative, three-dimensional over their left nucleus, and six-dimensional over their centre have also been classified for odd [40] and even characteristic [45].

4.4 The Tensor Rank of a Semifield

We saw in Section 4.1.2 that semifields can be represented by threefold tensors. We saw also that the space of tensors has a natural function called the *tensor rank*, defined with respect to the pure tensors (the Segre variety).

- What does the tensor rank tell us about the semifield?
- Can we actually calculate the tensor rank?

It turns out that the tensor rank measures the *multplicative complexity* of the semifield. Consider for example the problem of efficiently multplying two 2×2 matrices. Naively, we can do this by performing eight field multiplications (and some additions, which are regarded as free). However, due to Strassen's famous algorithm, we can actually reduce this to just seven multiplications. This is because the tensor rank of the tensor corresponding to matrix multiplication is seven. The question for 3×3 matrices is still open; the tensor rank lies somewhere between 19 and 23.

The problem of determining the tensor rank of a finite field extension $\mathbb{F}_{q^n}$ of $\mathbb{F}_q$ is a very important one for practical applications. When q is large enough with respect to n, polynomial interpolation can be used to achieve the minimum possible rank, namely $2n - 1$.

However in practise most relevant computations are done over extensions of $\mathbb{F}_2$. The best known algorithms use constructions from algebraic curves via modifications of the Chudnovsky-Chudnovsky algorithm [11].

The following table lists the best known lower bounds for the tensor rank of any semifields n-dimensional over $\mathbb{F}_q$, and upper bounds for the tensor rank of the field $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$.

	$q=2$		$q=3$	
n	LB	UB	LB	UB
2	3	3	3	3
3	6	6	6	6
4	9	9	8	9
5	13	13	10	12
6	15	15	12	15
7	18	22	17	19
8	20	24	19	21
9	26	30	21	26
10	28	33	24/25	27

Recently in [37], Michel Lavrauw and I demonstrated the first known example of a semifield having lower tensor rank than the field of the same order. In fact, we showed that the field and twisted field of order 3^4 have tensor rank nine, whereas the remaining semifields of this orde all have tensor rank eight. This suggests that some semifields could in theory have a computational efficiency advangate over the field of the same order.

It remains to explore cases where this may be possible. The case of semifields of order 2^8 is intriguing; the field $\mathbb{F}_{2^8}$ is used in many applicaitions, and there is plenty of room for a semifield of this order to have significanlty lower tensor rank.

References

[1] A. A. Albert. Generalized twisted fields. *Pacific J. Math.*, 11:1–8, 1961.

[2] K. Allen and J. Sheekey. On translation hyperovals in semifield planes, 2023.

[3] J. André. Über nicht-Desarguessche Ebenen mit transitiver Translationsgruppe. *Math. Z.*, 60:156–186, 1954.

[4] S. Ball, A. Blokhuis, and M. Lavrauw. Linear $(q+1)$-fold blocking sets in $\mathrm{PG}(2, q^4)$. *Finite Fields Appl.*, 6(4):294–301, 2000.

[5] S. Ball, G. Ebert, and M. Lavrauw. A geometric construction of finite semifields. *J. Algebra*, 311(1):117–129, 2007.

[6] D. Bartoli, B. Csajbók, G. Marino, and R. Trombetti. Evasive subspaces. *Journal of Combinatorial Designs*, 29(8):533–551, 2021.

[7] A. Blokhuis and M. Lavrauw. Scattered spaces with respect to a spread in $\mathrm{PG}(n, q)$. *Geom. Dedicata*, 81(1-3):231–243, 2000.

[8] A. Blokhuis, M. Lavrauw, and S. Ball. On the classification of semifield flocks. *Adv. Math.*, 180(1):104–111, 2003.

[9] R. H. Bruck and R. C. Bose. The construction of translation planes from projective spaces. *J. Algebra*, 1:85–102, 1964.

[10] I. Cardinali, O. Polverino, and R. Trombetti. Semifield planes of order q^4 with kernel F_{q^2} and center F_q. *European J. Combin.*, 27(6):940–961, 2006.

[11] D. V. Chudnovsky and G. V. Chudnovsky. Algebraic complexities and algebraic curves over finite fields. *J. Complexity*, 4(4):285–316, 1988.

[12] L. E. Dickson. On finite algebras. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1905:358–393, 1905.

[13] L. E. Dickson. On commutative linear algebras in which division is always uniquely possible. *Trans. Amer. Math. Soc.*, 7(4):514–522, 1906.

[14] E. F. Combarro, I. F. Rúa, and J. Ranilla. New advances in the computational exploration of semifields. *Int. J. Comput. Math.*, 88(9):1990–2000, 2011.

[15] E. F. Combarro, I. F. Rúa, and J. Ranilla. Finite semifields with 7^4 elements. *Int. J. Comput. Math.*, 89(13-14):1865–1878, 2012.

[16] J. de la Cruz, M. Kiermaier, A. Wassermann, and W. Willems. Algebraic structures of MRD codes. *Adv. Math. Commun.*, 10(3):499–510, 2016.

[17] P. Delsarte. Bilinear forms over a finite field, with applications to coding theory. *J. Combin. Theory Ser. A*, 25(3):226–241, 1978.

[18] U. Dempwolff. Semifield planes of order 81. *J. Geom.*, 89(1-2):1–16, 2008.

[19] U. Dempwolff. Autotopism groups of cyclic semifield planes. *J. Algebraic Combin.*, 34(4):641–669, 2011.

[20] M. J. Ganley and V. Jha. On a conjecture of Kallaher and Liebler. *Geom. Dedicata*, 21(3):277–289, 1986.

[21] F. Göloğlu and L. Kölsch. An exponential bound on the number of non-isotopic commutative semifields. *Trans. Amer. Math. Soc.*, 376(3):1683–1716, 2023.

[22] R. Gow and J. Sheekey. On primitive elements in finite semifields. *Finite Fields Appl.*, 17(2):194–204, 2011.

[23] A. Gruica, A. Ravagnani, J. Sheekey, and F. Zullo. Generalised scattered subspaces. *CoRR*, abs/2207.01027, 2022.

[24] D. R. Hughes and E. Kleinfeld. Seminuclear extensions of Galois fields. *Amer. J. Math.*, 82:389–392, 1960.

[25] V. Jha and N. L. Johnson. An analog of the Albert-Knuth theorem on the orders of finite semifields, and a complete solution to Cofman's subplane problem. *Algebras Groups Geom.*, 6(1):1–35, 1989.

[26] V. Jha, O. Polverino, and R. Trombetti. Subplanes of a translation plane. *Finite Fields Appl.*, 30:121–138, 2014.

[27] N. L. Johnson, G. Marino, O. Polverino, and R. Trombetti. Semifields of order q^6 with left nucleus $\mathbb{F}_{q^3}$ and center $\mathbb{F}_q$. *Finite Fields Appl.*, 14(2):456–469, 2008.

[28] M. J. Kallaher and R. A. Liebler. A conjecture on semifield planes. II. *Geom. Dedicata*, 8(1):13–30, 1979.

[29] W. M. Kantor. Commutative semifields and symplectic spreads. *J. Algebra*, 270(1):96–114, 2003.

[30] W. M. Kantor and R. A. Liebler. Semifields arising from irreducible semilinear transformations. *J. Aust. Math. Soc.*, 85(3):333–339, 2008.

[31] D. E. Knuth. Finite semifields and projective planes. *J. Algebra*, 2:182–217, 1965.

[32] O. V. Kravtsova. On alternating subgroup a_5 in autotopism group of finite semifield plane. 2020.

[33] M. Lavrauw. Sublines of prime order contained in the set of internal points of a conic. *Des. Codes Cryptogr.*, 38(1):113–123, 2006.

[34] M. Lavrauw and O. Polverino. Finite semifields. In L. Storme and J. De Beule, editor, *Current research topics in Galois geometry*, Mathematics Research Developments, pages 127–155. Nova Science, 2011.

[35] M. Lavrauw and J. Sheekey. Semifields from skew polynomial rings. *Adv. Geom.*, 13(4):583–604, 2013.

[36] M. Lavrauw and J. Sheekey. The BEL-rank of finite semifields. *Des. Codes Cryptogr.*, 84(3):345–358, 2017.

[37] M. Lavrauw and J. Sheekey. The tensor rank of semifields of order 16 and 81. *Linear Algebra Appl.*, 643:99–124, 2022.

[38] G. Lunardon, G. Marino, O. Polverino, and R. Trombetti. Translation dual of a semifield. *J. Combin. Theory Ser. A*, 115(8):1321–1332, 2008.

[39] G. Marino and O. Polverino. On the nuclei of a finite semifield. In *Theory and applications of finite fields*, volume 579 of *Contemp. Math.*, pages 123–141. Amer. Math. Soc., Providence, RI, 2012.

[40] G. Marino and V. Pepe. On symplectic semifield spreads of $\mathrm{PG}(5, q^2)$, q odd. *Forum Math.*, 30(2):497–512, 2018.

[41] G. Marino, O. Polverino, and R. Trombetti. On $\mathbb{F}_q$-linear sets of $\mathrm{PG}(3, q^3)$ and semifields. *J. Combin. Theory Ser. A*, 114(5):769–788, 2007.

[42] R. McConnel. Pseudo-ordered polynomials over a finite field. *Acta Arith.*, 8:127–151, 1962/63.

[43] G. Menichetti. On a Kaplansky conjecture concerning three-dimensional division algebras over a finite field. *J. Algebra*, 47(2):400–410, 1977.

[44] G. Menichetti. n-dimensional algebras over a field with a cyclic extension of degree n. *Geom. Dedicata*, 63(1):69–94, 1996.

[45] V. Pepe. Symplectic semifield spreads of $\mathrm{PG}(5, q^t)$, q even. *Ars Math. Contemp.*, 17(2):515–524, 2019.

[46] J.-C. Petit. Quasi-corps généralisant un type d'anneau quotient. *C. R. Acad. Sci. Paris Sér. A-B*, 265:A708–A711, 1967.

[47] I. F. Rúa. Primitive and non primitive finite semifields. *Comm. Algebra*, 32(2):793–803, 2004.

[48] I. F. Rúa. On the primitivity of four-dimensional finite semifields. *Finite Fields Appl.*, 33:212–229, 2015.

[49] I. F. Rúa, E. F. Combarro, and J. Ranilla. Classification of semifields of order 64. *J. Algebra*, 322(11):4011–4029, 2009.

[50] I. F. Rúa, E. F. Combarro, and J. Ranilla. Determination of division algebras with 243 elements. *Finite Fields Appl.*, 18(6):1148–1155, 2012.

[51] J. Sheekey. New semifields and new MRD codes from skew polynomial rings. *J. Lond. Math. Soc. (2)*, 101(1):432–456, 2020.

[52] R. J. Walker. Determination of division algebras with 32 elements. In *Proc. Sympos. Appl. Math., Vol. XV*, pages 83–85. Amer. Math. Soc., Providence, RI, 1963.

[53] Z.-X. Wan. *Geometry of matrices*. World Scientific Publishing Co., Inc., River Edge, NJ, 1996. In memory of Professor L. K. Hua (1910–1985).