CRYPTOGRAPHY MEETS FINITE GEOMETRY

Maxime Deryck

Promotors:

prof. dr. Jan De Beule *(Vrije Universiteit Brussel)*
prof. dr. Leo Storme *(Universiteit Gent)*

Master's thesis submitted in partial fulfilment
of the requirements for obtaining the academic degree of
Master of Science in Mathematics.

Academic year 2025–2026

Ghent University
Faculty of Sciences
Department of Mathematics:
Analysis, Logic and Discrete Mathematics

Cryptography meets Finite Geometry

Maxime Deryck

GHENT UNIVERSITY

Academic year 2025–2026

Promotors:

prof. dr. Jan De Beule
(Vrije Universiteit Brussel)

prof. dr. Leo Storme
(Universiteit Gent)

Master's thesis submitted in partial fulfilment of the requirements for obtaining the academic degree of Master of Science in Mathematics.

Preface

> *Schrijven is schrappen.*
>
> – Godfried Bomans

Reading research articles and writing my master's thesis is something I have genuinely looked forward to throughout my studies. Now that I have almost completed this task and am wrapping up my degree, I still consider it the most enjoyable part of my academic journey, especially since I was also able to discover some new results. A large part of why I liked it so much is, of course, the subject matter itself: cryptography viewed through a geometric lens. Having been fascinated by classical ciphers for a long time, and gradually growing more captivated by the complex mathematics underlying modern cryptography, I found it incredibly rewarding to take a deep dive into this theory for my thesis. The geometric perspective made it even more special, as it is a domain I had not studied much before.

For the opportunity to work on this fascinating subject, I must thank my promotors. First of all, I wish to thank prof. dr. Jan De Beule. Not only did he demonstrate how pure mathematics can have beautiful applications through the "Computeralgebra" course – this made me realise I was not quite made for functional analysis, another potential thesis subject – but he also took me on to work on code-based cryptography and welcomed me for enjoyable meetings at the VUB. Secondly, I would like to thank prof. dr. Leo Storme, whom I have known for a bit longer, as, in fact, my very first lecture at Ghent University was taught by him. His passionate teaching is clearly reflected in the support he provided during my work on this thesis: proposing papers tailored to my interests, suggesting research directions, and extensively proofreading my drafts.

That all this proofreading was no trivial task is evident from the page count. Although the reader might find it hard to believe, I still felt compelled to omit a substantial amount of content. This is precisely why I chose the specific quote on writing at the top of this page (loosely translated: to write is to edit); the hard part of writing really is the editing.

I should perhaps apologise to my girlfriend for occasionally portraying this as a Herculean task, especially since I did not face a strict word limit like she did while we worked on our theses simultaneously. Despite being much busier with lab work and internships, she always allowed me to share my enthusiasm about my thesis, for which I am deeply thankful. The same gratitude extends to my parents and sister, who have patiently listened to a great deal of mathematics that they probably did not quite understand.

Finally, I would like to thank you, the reader, for your interest in my thesis. I hope you will enjoy the remainder of this work.

Maxime Deryck
May 2026

Contents

List of symbols

We list some notations which will be repeatedly used in this thesis.

- $\mathbb{N}$ denotes the set of natural numbers, which includes 0;
- $\mathbb{F}_q$ denotes the finite field with $q = p^h$ elements, where p is prime and h a non-negative integer;
- $\mathrm{Gal}(E/F)$ denotes the Galois group, i.e. the automorphism group, of the field extension E/F over F;
- $\mathbb{F}_q^{m\times n}$ denotes the set of all $m \times n$ matrices over $\mathbb{F}_q$;
- $V(k, q)$ and $\mathbb{F}_q^k$ both denote the k-dimensional vector space over $\mathbb{F}_q$;
- $\mathrm{PG}(k-1, q)$ and $\mathrm{PG}(V, \mathbb{F}_q)$ both denote the $(k-1)$-dimensional projective space corresponding to the vector space $V = V(k, q)$ over $\mathbb{F}_q$;
- $\mathrm{Hom}_{\mathbb{F}_q}(V, W)$ denotes the space of all $\mathbb{F}_q$-linear maps (or $\mathbb{F}_q$-homomorphisms) from $V = V(m, q)$ to $W = V(n, q)$;
- $\mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_q^n)$ denotes the space of all $\mathbb{F}_q$-linear maps from $\mathbb{F}_q^n$ to itself (or $\mathbb{F}_q$-endomorphisms);
- $\mathrm{Aut}(\mathbb{F}_q)$ denotes the group of all automorphisms of $\mathbb{F}_q$;
- $\Gamma\mathrm{L}_n(\mathbb{F}_q)$ denotes the general semilinear group of degree n over $\mathbb{F}_q$, consisting of all invertible semilinear maps of $\mathbb{F}_q^n$;
- $\mathrm{P}\Gamma\mathrm{L}_n(\mathbb{F}_q)$ denotes the projective semilinear group of degree n over $\mathbb{F}_q$, consisting of all invertible semilinear maps of $\mathbb{F}_q^n$ modulo scalar multiples;
- $\mathrm{GL}_n(\mathbb{F}_q)$ denotes the general linear group of degree n over $\mathbb{F}_q$, consisting of all $n \times n$ invertible matrices with entries in $\mathbb{F}_q$;
- $\mathrm{PGL}_n(\mathbb{F}_q)$ denotes the projective general linear group of degree n over $\mathbb{F}_q$, consisting of all $n \times n$ invertible matrices with entries in $\mathbb{F}_q$ modulo scalar multiples;
- $A^\top$ (resp. $\mathbf{v}^\top$) denotes the transpose matrix (resp. vector) of $A \in \mathbb{F}_q^{m\times n}$ (resp. $\mathbf{v} \in \mathbb{F}_q^k$);
- $\mathrm{tr}(A)$ denotes the matrix trace of $A \in \mathbb{F}_q^{m\times n}$.

Throughout this text, we treat vectors $\mathbf{v} \in V(k, q)$ as row vectors. Consequently, linear maps are represented by matrices acting on the right (i.e. $\mathbf{v} \mapsto \mathbf{v}M$). In other words, we work with row spaces.

1 Introduction

> *Tant que l'Algèbre et la Géométrie ont été séparées, leurs progrès ont été lents et leurs usages bornés; mais lorsque ces deux sciences se sont réunies, elles se sont prêté des forces mutuelles et ont marché ensemble d'un pas rapide vers la perfection.*
>
> – Joseph-Louis Lagrange

Cryptography is a cornerstone of the modern digital world, integrated into many daily interactions. It is encountered, among other instances, when browsing the web, using payment terminals to authenticate transactions via digital signatures, and accessing digital broadcasts (television) through authentication protocols. By offering guarantees of confidentiality, integrity, and authentication, cryptography provides the trust needed for our ever-increasing reliance on digital infrastructure.

While there are many cryptographic disciplines, they are generally divided into three branches: symmetric (secret-key) cryptography, asymmetric (public-key) cryptography, and hash functions. Simply put, symmetric cryptography relies on a single shared key to both encrypt and decrypt data. In contrast, asymmetric cryptography uses a pair of keys: a public key that anyone can use to encrypt a message, and a private key kept secret by the recipient to decrypt it. Hash functions are primarily concerned with data integrity and will not be explored further in this thesis. All of these areas are driven by the perpetual cat-and-mouse game between cryptographers designing systems and cryptanalysts seeking to break them; as a result, research in the field is constantly advancing.

Regarding asymmetric cryptography in particular, we are currently at a crossroads. Historically, the security of widely deployed asymmetric cryptosystems has relied on the intractability of specific number-theoretic problems. However, the expected arrival of large-scale quantum computers, capable of breaking such systems rapidly, poses a critical threat to this infrastructure. Because adversaries can intercept and store encrypted data today with the intention of decrypting it once quantum hardware is realised, governments and businesses are forced to transition to post-quantum cryptography (PQC) as soon as possible. While some believe it will take many decades for these computers to be functional, others suggest they could be operational within a few years. Symmetric cryptosystems are notably more resilient to the quantum threat, yet they are not immune to progress; attacks against them also evolve constantly, meaning that rigorous analysis remains essential to refine current systems or design more robust alternatives.

This thesis investigates specific topics within these two domains. Accordingly, the text is divided into two main parts, which can be read independently of each other:

- **Part I: Code-based cryptography** focuses on asymmetric post-quantum cryptosystems based on hard problems from coding theory, directly motivated by the quantum computing threat.
- **Part II: Symmetric cryptography** studies the security properties of the building blocks used in symmetric ciphers.

In both parts of this thesis, a fruitful interplay between cryptography and finite geometry is demonstrated; see Figure 1.1. Broadly speaking, finite geometry is the study of geometric systems that contain a finite number of points. It explores discrete incidence structures, such as projective and affine spaces defined over finite fields, and investigates the properties of geometric objects within them. The ongoing research in this field is driven by its ability to bridge abstract algebra and combinatorics through geometric intuition. By translating algebraic constraints into geometric configurations, such as points, lines,

and intersecting subspaces, we can solve problems that are otherwise difficult to tackle. This geometric lens often reveals hidden structural properties, simplifies proofs, and yields novel construction methods that might remain entirely obscured in a strictly algebraic setting.

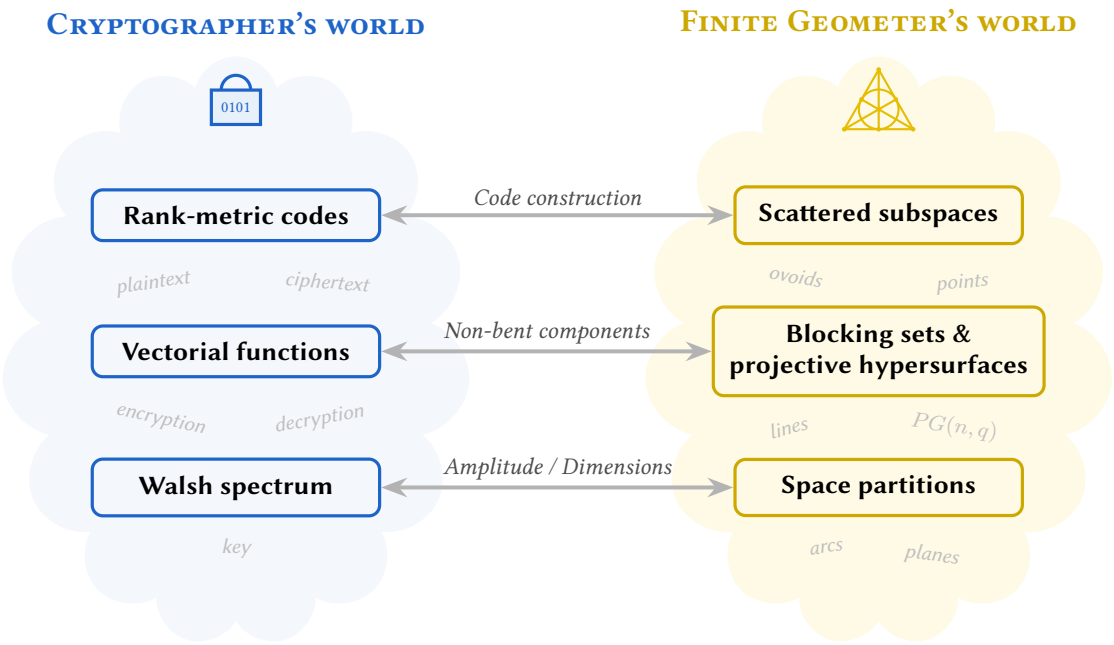


Figure 1.1: The interplay between cryptography and finite geometry explored in this thesis.

Within the landscape of PQC, code-based cryptography is a promising candidate for standardisation efforts. The fundamental mechanism of such cryptosystems relies on the intentional introduction of errors into a message encoded with an error-correcting code. Legitimate users possess the algebraic structure required for efficient error correction, whereas adversaries face the generally hard problem of decoding what appears to be a random (unstructured) code.

A significant limitation of classical code-based cryptosystems, which operate in the Hamming metric, is the requirement for very large public keys to maintain adequate security levels against generic decoding attacks. The rank metric was proposed as a solution to this inefficiency. Because the search space for rank errors scales exponentially with the field size, generic decoding in the rank metric is considerably harder than in the Hamming metric for equivalent parameters. This property allows rank-metric cryptosystems to achieve comparable security with significantly shorter code lengths and thus smaller public keys. Of particular interest are maximum-rank distance (MRD) codes that achieve the highest possible error-correcting capability for given code parameters.

For decades, Gabidulin codes [21] were the only known family of MRD codes, serving as the foundation for the GPT cryptosystem in 1991 [29]. However, the highly structured algebraic nature of Gabidulin codes, while necessary for efficient decoding, introduces critical vulnerabilities. Structural attacks, such as those found in [35, 77, 81], exploit these vulnerabilities to reconstruct an equivalent private key directly from the public key. Despite numerous proposals attempting to mask the Gabidulin structure via distortion and scrambling matrices, these techniques were consistently proven insufficient, rendering the original GPT cryptosystem and its direct successors insecure.

This demonstrated that secure rank-metric cryptography requires alternative (MRD) codes. The discovery of these new codes is largely driven by a recent connection to finite geometry [94]. By translating the algebraic constraints of MRD codes into geometric configurations, specifically through the study of scattered subspaces and linear sets [96, 104], researchers have successfully constructed novel families of (inequivalent) MRD codes.

The objective of the first part of the thesis is to provide a comprehensive, structured overview of

this domain. The corresponding chapters detail the theoretical foundations of rank-metric codes, trace the historical cryptanalysis of the GPT cryptosystem, and examine the geometric constructions that currently drive the search for new MRD codes.

In contrast, the second part of this thesis presents novel results. This part shifts the focus to the security properties of the fundamental building blocks used in symmetric block ciphers. Specifically, it investigates vectorial functions $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$, which frequently act as substitution boxes (S-boxes) to confuse the relationship between the ciphertext and the secret key in symmetric cryptosystems. A crucial property of a cryptographically secure vectorial function is high non-linearity. This resilience is quantified by the Walsh spectrum, which measures how closely a vectorial function can be approximated by a simple affine equation.

While the application of finite geometry to code-based cryptography in Part I is well-established, its application to symmetric cryptography in Part II is perhaps more surprising. In recent work, Bénéteau et al. [7] explored the Walsh spectrum of quadratic almost perfect nonlinear (APN) functions in characteristic two. They established a connection between these cryptographic functions and structures in finite geometry. Specifically, they linked the function's non-bent components, i.e. the projections of the function that exhibit suboptimal non-linearity, to blocking sets in projective space. Furthermore, they connected the amplitude distribution, which signifies the number and severity of these weak spots, to vector space partitions. This geometric lens provides a powerful tool to limit the admissible configurations of these vectorial functions. By imposing strict structural bounds, we can formally show the inexistence of several hypothetical functions, ensuring that we do not search for cryptographically ideal structures that simply cannot exist.

In this thesis, we seek to build upon and significantly broaden the framework in [7]. Firstly, we extend the theoretical foundation to fields of odd characteristic using the class of crooked quadratic functions. We demonstrate that the structural link to vector space partitions is maintained across different characteristics, allowing us to generalise existing combinatorial bounds on the amplitude distribution. Crucially, we introduce a novel algebraic perspective: we characterise the underlying structure of the set of non-bent components as a projective hypersurface.

In the binary case, this approach resolves an open problem originally posed in [7, Open Problem 6.4] regarding the minimum size of the set of non-bent components. We prove that their proposed lower bound cannot be reached by any quadratic APN function (a subset of non-MNBC quadratic functions in this context).

Theorem 5.4.10. *Let n be even with $n \leq 2m-6$. There does* not *exist a non-MNBC quadratic function $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ with $|N_F| = 2^{m-n/2} + 2^{m-n/2-2} + 2^{m-n/2-3} - 1$.*

Following this negative result, we leverage the weight distribution of Reed-Muller codes, connected via the polynomial defining our hypersurface, to establish a new and demonstrably sharp lower bound. This immediately yields an upper bound on the number of bent components for quadratic APN functions, which theoretically limits the cryptographic security of such functions.

Corollary 5.4.17. *Let $n \geq 4$ be even and $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a quadratic APN function. Then*

$$|N_F| \geq 2^{n/2} + 2^{n/2-1} - 1.$$

In other words, the number of bent component functions of F is less than or equal to $2^n - 2^{n/2} + 2^{n/2-1}$.

Furthermore, this geometric framework applies independently of the characteristic. In odd characteristic, the projective hypersurface dictates geometric properties regarding the intersection of the set of non-bent components with projective lines in the dual space. To the best of our knowledge, this is the first result of its kind.

Theorem 5.4.18. *Let $p > n \geq 2$ and let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a quadratic function. Consider a line $\mathcal{L} \subseteq \mathrm{PG}(n-1,p)^*$. Let $k = |\mathcal{L} \cap N_F|$ be the number of non-bent components on this line. Then k must satisfy one of the following:*

1. $k \leq n$,
2. $k = p + 1$ *(the entire line is contained in* N_F*).*

Ultimately, the second part of this thesis highlights how translating algebraic and cryptographic properties into finite geometry not only simplifies their analysis but also unlocks stronger bounds on the security of symmetric ciphers.

Outline. We end this introduction with an overview of what the reader can expect from this work. For a Dutch summary, see Appendix A.

Part I is dedicated to code-based cryptography. In Chapter 2, the foundations of rank-metric codes are introduced alongside the original GPT cryptosystem. Chapter 3 provides an in-depth study of maximum-rank distance (MRD) codes, exploring both their algebraic representations and the geometric equivalences used to construct new code families. The necessity for these novel geometric constructions is subsequently demonstrated in Chapter 4, which details the historical structural cryptanalysis that ultimately compromised the classical GPT cryptosystem.

Part II contains the novel research of this thesis within the domain of symmetric cryptography. Chapter 5 investigates the Walsh spectrum of cryptographic functions. It first reviews the known connections to vector space partitions and blocking sets in characteristic two, before detailing our generalisation to odd characteristic and the strict theoretical bounds derived from our novel geometric characterisation.

Finally, this thesis ends with Chapter 6, which provides a general conclusion of this work and outlines several directions for future research.

As demonstrated by the formatting of this paragraph, original contributions (to the best of our knowledge) within this thesis are distinguished by a solid yellow line in the left margin alongside the theorem statement. This applies to both entirely novel results as well as straightforward extensions of previous work. To clarify the attribution of proofs throughout the text, the following system is employed. A blue square (■) indicates that the proof has been reproduced or adapted from the cited source. Conversely, a yellow square (■) denotes a self-constructed proof. This encompasses proofs of original contributions, proofs omitted from their cited sources, and proofs of mathematical folklore where an explicit reference is absent but the formal argument has been independently formulated to fit the narrative.

Part I

Code-based cryptography

2 Foundations of code-based cryptography

In coding theory, the primary goal is to detect and correct errors that occur naturally over a noisy channel. By correcting these errors, a distorted message can be interpreted correctly by the receiver. To decode the received message, it is necessary to know the code parameters and possess an efficient decoding algorithm. In 1978, Robert McEliece realised that deliberately introducing errors into a message, while keeping the code parameters private, could be used for cryptography [68]. Indeed, an adversary who does not possess the specific information required for efficient decoding cannot recover the original message. Consequently, the confidentiality of the transmission is ensured.

Since then, code-based cryptography (CBC) has been extensively studied. Due to recent standardisation efforts by the United States' National Institute of Standards and Technology (NIST) for post-quantum cryptosystems, research in this field has intensified. To this day, code-based cryptography is regarded as quantum-secure. In March 2025, a code-based key encapsulation mechanism (KEM) – used to share keys securely between parties over an insecure channel – was selected for standardisation by NIST. This will complement previously selected systems, providing an alternative should security weaknesses be discovered in other approaches. Furthermore, two code-based schemes reached the second round of the call for (additional) digital signature schemes, although neither proceeded to the next round.

This thesis does not focus on those specific proposals, nor does it provide an exhaustive overview of code-based cryptography; for a comprehensive survey, see [103]. Instead, the first part of this thesis focuses on rank-metric code constructions designed to address the limitations of existing code families within the context of code-based cryptography. In particular, this chapter establishes the concepts required to understand the foundations of code-based cryptography. It specifically examines the original McEliece framework and its underlying security assumptions. Finally, the GPT cryptosystem is introduced, serving as the primary motivation and focus of the subsequent chapters.

2.1 Relevant concepts from coding theory

In classical coding theory, a (block) code $\mathcal{C}$ is simply any subset $\mathcal{C} \subseteq \mathbb{F}_q^n$ of the n-dimensional vector space $\mathbb{F}_q^n$ over $\mathbb{F}_q$.

Definition 2.1.1 (Linear code). Let $1 \leq k \leq n$ be integers. A *linear* $[n, k]$*-code*, or simply $[n, k]$-code, over $\mathbb{F}_q$ is a k-dimensional linear subspace of $\mathbb{F}_q^n$.

One of the advantages of linear codes is that they can be represented easily and compactly through a generator matrix.

Definition 2.1.2 (Generator matrix). Let $k \leq n$ be positive integers and let $\mathcal{C}$ be a linear $[n, k]$-code over $\mathbb{F}_q$. Then, a matrix $G \in \mathbb{F}_q^{k \times n}$ is called a *generator matrix* of $\mathcal{C}$ if

$$\mathcal{C} = \left\{ \mathbf{x}G : \mathbf{x} \in \mathbb{F}_q^k \right\},$$

that is, the rows of G form a basis of $\mathcal{C}$.

In order to measure how far apart two vectors are, we can endow $\mathbb{F}_q^n$ with a metric.

Definition 2.1.3 (Hamming metric). For $\mathbf{x}, \mathbf{y} \in \mathbb{F}_q^n$, the *Hamming distance* between $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ is given by the number of positions in which they differ, i.e.

$$d_H(\mathbf{x}, \mathbf{y}) = |\{ i \in \{1, \ldots, n\} : x_i \neq y_i \}|.$$

The *Hamming weight* $\mathrm{wt}_H(\mathbf{x})$ of a vector $\mathbf{x} \in \mathbb{F}_q^n$ is the number of non-zero positions, or equivalently, $\mathrm{wt}_H(\mathbf{x}) = d_H(\mathbf{x}, \mathbf{0})$.

An important parameter of any code is the minimum distance, which is connected to the error-correcting capability of the code. In words, the minimum distance of a code is the smallest distance achieved by its distinct codewords.

Definition 2.1.4 (Minimum distance). Let $\mathcal{C}$ be a code endowed with a metric d_{metric}. The *minimum distance* of $\mathcal{C}$, denoted by $d(\mathcal{C})$, is given by

$$d(\mathcal{C}) = \min\{d_{\text{metric}}(c, c') : c, c' \in \mathcal{C}, c \neq c'\}.$$

Notation 2.1.5. When the minimum distance d of an $[n, k]$-code $\mathcal{C}$ is known, it is often included in the parameters, and we call $\mathcal{C}$ an $[n, k, d]$-code.

In the Hamming metric, a variety of bounds exist that establish the theoretical limits of code performance; see, for example, [43]. One such fundamental result is the Singleton bound [99]. It famously provides an upper bound on the maximum number of possible codewords, denoted by $A_q(n, d)$, in a q-ary code of length n with minimum distance d.

Theorem 2.1.6 ([99, Theorem 1]). *For a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ with minimum distance d in the Hamming metric, we have*

$$|\mathcal{C}| \leq q^{n-d+1},$$

whence $A_q(n, d) \leq q^{n-d+1}$. Furthermore, if $\mathcal{C}$ is a linear $[n, k, d]$-code over $\mathbb{F}_q$, then

$$d \leq n - k + 1.$$

A Hamming metric code that achieves the Singleton bound is called a *maximum distance separable* (MDS) code. For fixed parameters, these codes possess the greatest error-detection and error-correction capabilities.

While codes in the Hamming metric have been studied most intensively, there exist many more metrics. Of particular interest for this thesis are rank-metric codes, for which the most general definition is given next.

Definition 2.1.7 (Rank-metric code). Let $\mathbb{F}_q^m$ and $\mathbb{F}_q^n$ be vector spaces of dimension m and n over $\mathbb{F}_q$, respectively. A *rank-metric code* $\mathcal{C}$ is a non-empty subset of $\mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$ endowed with the rank metric d_R, defined by

$$d_R(f, g) = \mathrm{rank}(f - g) = \dim(\mathrm{Im}(f - g)),$$

for all $f, g \in \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$. If $\mathcal{C}$ is a linear subspace of $\mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$, it is called an $\mathbb{F}_q$*-linear* (m, n, q) *rank-metric code*, or simply (m, n, q)-code.

Notation 2.1.8. When the minimum rank distance d of an (m, n, q)-code $\mathcal{C}$ is known, it is often included in the parameters, and we call $\mathcal{C}$ an $(m, n, q; d)$-code.

Remark 2.1.9. In the theory of rank-metric codes, it is important to distinguish between linearity over the base field $\mathbb{F}_p$ and linearity over an extension field $\mathbb{F}_q$, with $q = p^h$ and $h \geq 2$. To avoid ambiguity, we will explicitly specify the field of linearity, for instance, by writing "$\mathbb{F}_p$-linear" or "$\mathbb{F}_q$-subspace".

A Singleton-like bound also exists for rank-metric codes, as first proved by Delsarte [21].

Theorem 2.1.10 ([21, Theorem 5.4]). *Let the rank-metric code $\mathcal{C} \subseteq \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$ have minimum rank distance d. Then*

$$|\mathcal{C}| \leq q^{\max\{m,n\}(\min\{m,n\}-d+1)}. \tag{2.1}$$

Furthermore, if $\mathcal{C}$ is $\mathbb{F}_q$-linear with $k = \dim_{\mathbb{F}_q}(\mathcal{C})$, then

$$k \leq \max\{m, n\}(\min\{m, n\} - d + 1). \tag{2.2}$$

Proof. Suppose without loss of generality that $n \geq m$. Assume, to the contrary, that $|\mathcal{C}| > q^{n(m-d+1)}$. Let $\varphi \colon \mathbb{F}_q^{m-d+1} \hookrightarrow \mathbb{F}_q^m$ be any injective linear map. Define

$$\Phi \colon \mathcal{C} \to \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^{m-d+1}, \mathbb{F}_q^n), f \mapsto \Phi(f) = f \circ \varphi.$$

The codomain of Φ has at most $q^{n(m-d+1)}$ elements. Therefore, by the pigeonhole principle, there exist distinct $f, g \in \mathcal{C}$ with $\Phi(f) = \Phi(g)$, i.e. $(f - g) \circ \varphi = 0$. Thus, $\mathrm{Im}(\varphi) \subseteq \ker(f - g)$, so $\dim(\ker(f - g)) \geq m - d + 1$ and by the rank–nullity theorem

$$\mathrm{rank}(f - g) = m - \dim(\ker(f - g)) \leq d - 1 < d.$$

This is a contradiction with the minimum distance d. Therefore, $|\mathcal{C}| \leq q^{n(m-d+1)}$; the symmetric case $m \geq n$ is analogous and also yields the claimed bound. ∎

A rank-metric code that achieves the upper bound in the theorem above is called a *maximum-rank distance* (MRD) code.

2.2 Relevant concepts from cryptography

While coding theory enables *reliable* communication, cryptography is the art of *secure* communication. The cryptographic problem of interest in this part is situated within asymmetric cryptography, also known as public-key cryptography.

The concept of asymmetric cryptography was first introduced by Diffie and Hellman in 1976 [23]. In this setting, we distinguish between two entities: the sender, conventionally named Bob, and the receiver, named Alice. Unlike symmetric systems, where both parties must share a single secret key, asymmetric schemes use a pair of mathematically related keys: a public key and a private key. Alice generates this pair and shares the public key openly, while keeping the private key strictly secret. When Bob wishes to send a confidential message to Alice, he uses her public key to encrypt the data into a ciphertext. This ciphertext can be sent over an insecure channel because it is practically impossible for an attacker to decipher it without the private key. Upon reception, Alice uses her private key to decrypt the ciphertext and recover the original message.

For decades, digital security has relied on the difficulty of specific number-theoretic problems. The most famous examples are the RSA cryptosystem [90], based on the difficulty of factoring large integers, and elliptic curve cryptography (ECC), independently proposed by Koblitz [48] and Miller [70], which relies on the discrete logarithm problem and allows for smaller keys to obtain the same security level. However, in 1994, Peter Shor introduced a quantum algorithm capable of solving these problems efficiently [97]. Consequently, a large-scale quantum computer would render these classical systems insecure. This threat has led to the field of post-quantum cryptography (PQC), which seeks to find systems that remain secure against classical as well as quantum computers. Code-based cryptography, which does not rely on number theory, is considered a promising candidate for this purpose.

Fundamentally, the security of any cryptosystem rests on computational complexity theory. The distinction between the classes P and NP is of great importance. The class P contains problems that can be solved efficiently, that is, in polynomial time, while NP contains problems for which a proposed solution can be verified in polynomial time, even if finding such a solution may be much harder. Among the problems in NP, some are known to be the hardest: these are the NP-complete problems, defined as those in NP to which every other problem in NP can be reduced by a polynomial-time reduction. If any NP-complete problem could be solved in polynomial time, then all problems in NP could be, implying P = NP (which is widely believed not to be the case; see, for example, [25]). For cryptography, we rely on problems believed to lie outside of P: the intended operations should remain efficient for the users, but solving the underlying hard problem should be infeasible for an attacker.

2.3 The McEliece framework

Within the domain of code-based cryptography, two systems for asymmetric cryptography have historically dominated the field: the McEliece [68] and the Niederreiter [72] cryptosystems. Both cryptosystems are computationally equivalent [54]. However, as the McEliece framework provides the structural basis for the GPT cryptosystem (which is the main motivation for subsequent chapters), this section restricts its focus to McEliece's construction. Originally, this cryptosystem was introduced using binary Goppa codes; see, e.g. [8]. Nevertheless, the underlying ideas are not bound to a specific code family. The framework can be generalised to any family of codes that possesses an efficient decoding algorithm. Therefore, the following description does not impose a specific code.

The framework consists of three distinct phases: key generation, encryption and decryption.

Key generation. The receiver, Alice, performs the following steps to generate her public and private keys.

1. Alice selects a linear $[n, k]$-code $\mathcal{C}$ over $\mathbb{F}_q$ capable of efficiently correcting t errors using a decoding algorithm $\mathcal{D}$. Let $G \in \mathbb{F}_q^{k\times n}$ be a generator matrix for $\mathcal{C}$.

2. She selects a random non-singular matrix $S \in \mathrm{GL}_k(\mathbb{F}_q)$, often referred to as the scrambler matrix.

3. She selects a random permutation matrix $P \in \mathbb{F}_q^{n\times n}$.

4. Alice computes the public generator matrix $\hat{G} = SGP$.

The public key is the tuple $(\hat{G}, t)$. The private key is the tuple (S, G, P). The matrices S and P serve to disguise the structure of the code $\mathcal{C}$, making $\hat{G}$ appear as the generator matrix of a random linear code.

Encryption. To send a message to Alice, the sender, Bob, requires her public key $(\hat{G}, t)$.

1. Bob represents his message m as a vector $\mathbf{x} \in \mathbb{F}_q^k$.

2. He generates a random error vector $\mathbf{e} \in \mathbb{F}_q^n$ such that its Hamming weight satisfies $\mathrm{wt}_H(\mathbf{e}) \leq t$.

3. Bob computes the ciphertext $\mathbf{c} \in \mathbb{F}_q^n$ as

$$\mathbf{c} = \mathbf{x}\hat{G} + \mathbf{e}. \tag{2.3}$$

Once computed, the ciphertext $\mathbf{c}$ is transmitted to Alice over a (possibly) insecure channel.

Decryption. Upon receiving the ciphertext $\mathbf{c}$, Alice uses her private key (S, G, P) to recover the message.

1. Alice computes the inverse transformation by multiplying the ciphertext by P^{-1} (which is equal to P^T for permutation matrices):

$$\mathbf{c}' = \mathbf{c}P^{-1} \tag{2.4}$$

2. She applies the decoding algorithm $\mathcal{D}$ to $\mathbf{c}'$ to obtain a vector $\mathbf{x}'$.

3. Finally, Alice recovers the original message $\mathbf{x}'$ by computing $\mathbf{x} = \mathbf{x}'S^{-1}$.

The correctness of this recovery procedure relies on the the fact that P is an isometry (for the Hamming metric), which ensures the added error $\mathbf{e}$ remains decodable.

To verify that the decryption process yields the original message m, we substitute Equation (2.3) for the ciphertext $\mathbf{c}$ into Equation (2.4):

$$\begin{aligned}\mathbf{c}' &= \mathbf{c}P^{-1} \\ &= (\mathbf{x}\hat{G} + \mathbf{e})P^{-1} \\ &= (\mathbf{x}(SGP) + \mathbf{e})P^{-1} \\ &= \mathbf{x}SG(PP^{-1}) + \mathbf{e}P^{-1} \\ &= (\mathbf{x}S)G + \mathbf{e}P^{-1}.\end{aligned}$$

Let $\mathbf{x}' = \mathbf{x}S$ and $\mathbf{e}' = \mathbf{e}P^{-1}$. The term $\mathbf{x}'G$ represents a codeword in $\mathcal{C}$, as it is a linear combination of the rows of G. Since P is a permutation matrix, it acts as an isometry with respect to the Hamming metric; therefore, the Hamming weight is preserved, so $\mathrm{wt}_H(\mathbf{e}') = \mathrm{wt}_H(\mathbf{e}) \leq t$.

Because the error term $\mathbf{e}'$ has weight at most t, it falls within the error-correcting capability of the code $\mathcal{C}$. Therefore, the decoding algorithm $\mathcal{D}$ correctly identifies $\mathbf{x}'G$ and returns $\mathbf{x}' = \mathbf{x}S$. Alice then computes

$$\mathbf{x}'S^{-1} = (\mathbf{x}S)S^{-1} = \mathbf{x},$$

which confirms that the original message is successfully recovered.

Example 2.3.1. Consider the McEliece cryptosystem over the binary field $\mathbb{F}_2$. Alice selects a linear $[6, 3]$-code $\mathcal{C}$ with minimum (Hamming) distance $d = 3$, which can correct $t = 1$ error. She chooses the generator matrix G for $\mathcal{C}$ in systematic form, meaning the message is embedded in the first $k = 3$ positions of the codeword:

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

She also picks a non-singular scrambler matrix $S \in \mathrm{GL}_3(\mathbb{F}_2)$ and a permutation matrix $P \in \mathbb{F}_2^{6\times 6}$. Let S be a matrix that adds the second row to the first, and let P be a permutation that swaps the last two columns:

$$S = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Alice then computes the public key $\hat{G} = SGP$, yielding

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

The public key is $(\hat{G}, t = 1)$.

Bob wishes to send the message $\mathbf{x} = (1, 0, 1)$. He selects an error vector with weight $t = 1$, for instance, $\mathbf{e} = (0, 1, 0, 0, 0, 0)$. He computes the ciphertext:

$$\mathbf{c} = \mathbf{x}\hat{G} + \mathbf{e} = (1, 1, 1, 0, 0, 0) + (0, 1, 0, 0, 0, 0) = (1, 0, 1, 0, 0, 0).$$

Alice receives $\mathbf{c} = (1, 0, 1, 0, 0, 0)$. She computes $\mathbf{c}' = \mathbf{c}P^{-1}$. Since P swaps the last two columns, and $\mathbf{c}$ has zeros in those positions, the vector remains unchanged:

$$\mathbf{c}' = (1, 0, 1, 0, 0, 0).$$

Alice now decodes $\mathbf{c}'$ with respect to the original code $\mathcal{C}$. The error $\mathbf{e}P^{-1}$ is still at the second position. The decoding algorithm $\mathcal{D}$ corrects this error to recover the codeword $(1,1,1,0,0,0)$. Since G is systematic, the corresponding message vector $\mathbf{x}'$ consists of the first three bits: $\mathbf{x}' = (1,1,1)$.

Finally, Alice computes $\mathbf{x} = \mathbf{x}'S^{-1}$. The inverse of S (over $\mathbb{F}_2$) is identical to S itself:

$$\mathbf{x} = (1,1,1)\begin{pmatrix}1 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} = (1,0,1).$$

The original message is successfully recovered. □

The example above illustrates the mechanics of the system using trivial parameters. However, in a practical scenario, an adversary possessing only the public matrix $\hat{G}$ faces a formidable challenge. Unlike the small code used here, the codes employed in actual instances are of sufficiently high dimension that attempting to decode a general linear code becomes computationally infeasible. To understand the precise conditions under which this system remains secure against adversarial attacks, we must formalise the underlying hard problems and the specific security assumptions governing the McEliece framework.

2.4 Security assumptions

The security of the McEliece framework (and the equivalent Niederreiter framework) relies fundamentally on the intractability of two distinct computational problems. Consequently, the analysis of the system is typically divided into two categories: non-structural attacks and structural attacks, each corresponding to one of the computational hard problems.

Non-structural attacks do not exploit the specific algebraic structure of the code used in the key generation. Instead, they treat the public key as a random linear code and attempt to solve the underlying general decoding problem directly. The security against these attacks rests on the following assumption.

Assumption 1 (Hardness of decoding). Decoding a random linear code is hard/infeasible, i.e. given a random generator matrix G and a target vector $\mathbf{y} = \mathbf{x}G + \mathbf{e}$, it is computationally very hard to recover either $\mathbf{x}$ or $\mathbf{e}$.

Structural attacks, conversely, aim to invalidate the randomness of the public key. They attempt to exploit specific properties of the underlying code family (e.g. Goppa codes or Reed-Solomon codes) to recover the private structure (the secret matrices S, G, P). Resistance against these attacks rests on the indistinguishability assumption.

Assumption 2 (Indistinguishability). The public code is not distinguishable from a random code, i.e. the generator matrix $\hat{G}$ produced by the key generation algorithm is computationally indistinguishable from a generator matrix chosen uniformly at random from the set of all full-rank matrices over $\mathbb{F}_q$.

We now provide some further details regarding both types of attacks, reserving a more extensive analysis of structural attacks in the case of the GPT cryptosystem (see below) for Chapter 4.

2.4.1 Non-structural attacks

Non-structural attacks treat the public key as a random linear code, ignoring any potential underlying algebraic structure. The security against such attacks rests entirely on Assumption 1. In the literature, this assumption is typically formalised in two equivalent ways: the (computational) decoding problem (DP) and the syndrome decoding problem (SDP). The former concerns finding the message vector, while the latter is concerned with the error vector.

Definition 2.4.1 (Decoding problem). Let $\mathbb{F}_q$ be a finite field and $k \leq n$ be positive integers. Given $G \in \mathbb{F}_q^{k\times n}$, $\mathbf{r} \in \mathbb{F}_q^n$ and $t \in \mathbb{N}$, is there a vector $\mathbf{x} \in \mathbb{F}_q^k$, and a vector $\mathbf{e} \in \mathbb{F}_q^n$ of weight less than or equal to t such that $\mathbf{r} = \mathbf{x}G + \mathbf{e}$?

Definition 2.4.2 (Syndrome decoding problem). Let $\mathbb{F}_q$ be a finite field and $k \leq n$ be positive integers. Given $H \in \mathbb{F}_q^{(n-k)\times n}$, $\mathbf{s} \in \mathbb{F}_q^{n-k}$ and $t \in \mathbb{N}$, is there a vector $\mathbf{e} \in \mathbb{F}_q^n$ such that its weight is less than or equal to t and $\mathbf{e}H^T = \mathbf{s}$?

These two problems are computationally equivalent; see, for example, [103, Theorem 183]. To reduce a decoding instance to a syndrome decoding instance, one computes a parity-check matrix H from G and evaluates the syndrome $\mathbf{s} = \mathbf{r}H^\top$. Conversely, to translate a syndrome decoding instance into a decoding instance, one computes a generator matrix G from H and finds an arbitrary vector $\mathbf{r}$ satisfying $\mathbf{r}H^\top = \mathbf{s}$. Because both transformations rely entirely on polynomial-time linear algebra, an algorithm solving either problem can be readily adapted to solve the other.

The complexity of these problems depends on the metric used. In the Hamming metric, Berlekamp, McEliece, and van Tilborg proved that the syndrome decoding problem is NP-complete [9]. This proof provides the cornerstone for the security assumptions of Hamming-metric cryptosystems.

In the context of the GPT cryptosystem, which we consider later, the underlying code is equipped with the rank metric. The analogous problem is the *rank syndrome decoding problem* (Rank SDP). While the NP-completeness of the Rank SDP has not been proven directly, a probabilistic polynomial-time reduction from the SDP to the Rank SDP has been found [32]. This reduction provides a strong indication of the hardness of the Rank SDP because it succeeds "with high probability", but it remains an open question whether it really belongs to the class of NP-complete problems like the SDP.

Despite the probabilistic reduction to an NP-complete problem, which concerns worst-case complexity, practical security relies on the hardness of random instances. The most effective generic algorithms for solving the SDP in the Hamming metric are variants of *information set decoding* (ISD), first introduced by Prange [85]. The core idea of ISD is to search for an information set $I \subset \{1, \ldots, n\}$ of size k, defined such that the submatrix $\hat{G}_I$ formed by the columns of the public matrix $\hat{G}$ indexed by I is non-singular. The algorithm relies on the probability that such a set is free of errors. If the error vector $\mathbf{e}$ is zero at the positions in I (i.e. $\mathbf{e}_I = 0$), the encryption equation restricted to these coordinates simplifies:

$$\mathbf{c}_I = \mathbf{x}\hat{G}_I + \mathbf{e}_I \implies \mathbf{c}_I = \mathbf{x}\hat{G}_I.$$

Because $\hat{G}_I$ is invertible, the attacker can recover a candidate message via simple linear algebra:

$$\mathbf{x} = \mathbf{c}_I(\hat{G}_I)^{-1}.$$

This candidate $\mathbf{x}$ is then verified by re-encoding it to check if the distance between $\mathbf{c}$ and $\mathbf{x}\hat{G}$ is within the error-correcting capability t. Modern improvements to Prange's algorithm improve the search efficiency but still result in an exponential time complexity. Even quantum information set decoding algorithms [10, 24, 44] – while asymptotically faster than their classical counterparts – do not undermine Assumption 1, provided that the parameters (and therefore the key sizes) are chosen large enough to withstand these quantum improvements.

2.4.2 Structural attacks

While Assumption 1 ensures that a random code is hard to decode, the McEliece system does not use a random code; it uses a structured code masked by a scrambling and a permutation matrix. Security, therefore, holds only if Assumption 2 is satisfied. If this assumption fails, there exists a *distinguisher*, i.e. an algorithm that can differentiate the public matrix $\hat{G}$ from a random matrix. The existence of a distinguisher often implies that the specific code family leaks information, which can lead to full key recovery via a structural attack.

In a structural attack, the adversary seeks to reconstruct the private key (S, G, P) from the public key $\hat{G}$. This is typically achieved by identifying algebraic invariants of the code family that are preserved under the scrambling transformation. In Chapter 4, we explore these structural attacks further in the context of the GPT cryptosystem.

Thus, the choice of the underlying code family is critical. To date, binary Goppa codes, as utilised in the original McEliece proposal, remain one of the few families for which Assumption 2 is believed to hold, as no polynomial-time structural attack has been discovered despite decades of analysis [103].

2.4.3 Key sizes

Despite the strong security guarantees provided by the hardness of the decoding problem and the resistance of Goppa codes to structural attacks, the original McEliece cryptosystem has faced a significant hurdle regarding its adoption in practical applications: the size of the public key.

In the McEliece framework, the public key consists of the $k \times n$ generator matrix $\hat{G}$ over $\mathbb{F}_q$. Even when expressed in systematic form to reduce storage requirements, the size of the key grows quadratically with the code parameters. To resist information set decoding attacks (as discussed briefly in Section 2.4.1), the length n and dimension k must be chosen sufficiently large.

For example, achieving a security level of 128 bits – meaning that any attack is believed to require on the order of 2^{128} operations – with binary Goppa codes in "classic" McEliece requires public keys of roughly 192 to 260 kilobytes [11, 74]. In stark contrast, pre-quantum number-theoretic schemes used in practice, such as RSA-3072 or ECC-256, require keys of only a few kilobytes or mere bytes. While this discrepancy is less severe when compared to some other post-quantum candidates, it remains a critical bottleneck for resource-constrained devices and transmission bandwidth.

This efficiency drawback has driven a decades-long search for alternative code families that allow for smaller parameters without compromising security. The objective is to find codes where the structural masking (the indistinguishability assumption) holds, yet the data density allows for a more compact representation.

This motivation leads directly to the study of codes equipped with the rank metric, which will be the focus of the remainder of this part on code-based cryptography.

2.5 The GPT cryptosystem

The efficiency drawbacks of the original McEliece system, specifically the large public key sizes required to resist information set decoding, motivated the search for alternative code families. In 1991, Gabidulin, Paramonov, and Tretjakov proposed a new cryptosystem, now commonly referred to as the GPT cryptosystem [29].

The novelty of the GPT cryptosystem is the shift from the Hamming metric to the rank metric, as defined in Definition 2.1.7. The idea is to instantiate the McEliece framework using a family of rank-metric codes that possess an error-correcting capability significantly higher than random codes. In that regard, maximum-rank distance (MRD) codes are of particular interest because they achieve the highest error-correcting capability for given code parameters (length and dimension). In the rank metric, the error vector $\mathbf{e} \in \mathbb{F}_{q^m}^n$ is interpreted as follows: its "size" is measured by the rank of its matrix representation over the base field $\mathbb{F}_q$, rather than by the number of its non-zero positions. Since every $\mathbb{F}_q$-linear map admits a matrix representation, this point of view is natural in the rank metric. The specific representations, construction methods, and properties of rank-metric codes (MRD codes in particular) are the main focus of Chapter 3.

The workflow of the GPT cryptosystem – originally proposed using Gabidulin codes – mirrors the overall structure of the McEliece framework, but there are some differences due to the properties of the rank metric. Most notably, the permutation matrix P used in McEliece is ineffective here because the rank of a matrix is invariant under column permutations. Similarly, multiplying by a non-singular

scrambler matrix S over $\mathbb{F}_q$ merely represents a change of basis, meaning SG remains decodable by the same decoder. To effectively hide the code's structure, the GPT cryptosystem instead introduces a distortion matrix X.

Notation 2.5.1. To formally define the parameters of the GPT cryptosystem, it is crucial to distinguish between the rank over the base field $\mathbb{F}_q$ and the extension field $\mathbb{F}_{q^m}$. For a vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q^m}^n$, its rank over $\mathbb{F}_q$ is denoted by $\mathrm{rk}_q(\mathbf{x})$ and defined as the dimension of the $\mathbb{F}_q$-vector space spanned by its components: $\mathrm{rk}_q(\mathbf{x}) = \dim\left(\sum_{i=1}^{n} \mathbb{F}_q x_i\right)$. Analogously, for a matrix $M \in \mathbb{F}_{q^m}^{k\times n}$, its column rank over $\mathbb{F}_q$ is denoted by $\mathrm{rk}_q(M)$.

We now present the original GPT cryptosystem from [29] as outlined in [80], reserving later generalisations for the cryptanalysis discussion in Chapter 4. Moreover, the exact definition of a Gabidulin code is not required to understand its working at this point.

- **System parameters:** $q, n, m, k, \ell, s \in \mathbb{N}$, where $n \leq m$, $\ell < \frac{n-k-1}{2}$, and $s \leq \min\{k, \ell\}$.
- **Key generation:** First, generate the following matrices over $\mathbb{F}_{q^m}$:
  - G: $k \times n$ generator matrix of a $(k, n, q^m; d)$ Gabidulin code $\mathcal{G}$ with error-correcting capability t;
  - S: $k \times k$ random non-singular matrix (the row scrambler);
  - X: $k \times n$ random matrix with $\mathrm{rk}_{q^m}(X) = s$ and $\mathrm{rk}_q(X) = \ell$ (the distortion matrix).

  Then, compute the public generator matrix $\hat{G} = S(G + X)$ and error-correcting capability $\hat{t} = \lfloor \frac{n-k}{2} \rfloor - \ell$. Furthermore, let $\mathcal{D}_\mathcal{G}$ be an efficient decoding algorithm for $\mathcal{G}$.
- **Public key:** $(\hat{G}, \hat{t})$
- **Private key:** $(\mathcal{D}_\mathcal{G}, X, S)$ or (G, X, S)
- **Encryption:** To encode a plaintext $\mathbf{x} \in \mathbb{F}_{q^m}^k$, choose an error vector $\mathbf{e} \in \mathbb{F}_{q^m}^n$ with $\mathrm{rk}_q(\mathbf{e}) = \hat{t}$ at random and compute the ciphertext $\mathbf{c}$ as follows:

$$\mathbf{c} = \mathbf{x}\hat{G} + \mathbf{e}.$$

- **Decryption:** To decode a ciphertext $\mathbf{c}$, apply the decoding algorithm $\mathcal{D}_\mathcal{G}$ to it. Since $\mathbf{c}$ is at a rank distance less than or equal to $\lfloor \frac{n-k}{2} \rfloor$ from a codeword of $\mathcal{G}$, we obtain

$$\mathbf{x}SG = \mathcal{D}_\mathcal{G}(\mathbf{c}).$$

  From this, we can compute the plaintext $\mathbf{x}$.

A small example of this cryptosystem is given next.

Example 2.5.2. To illustrate a non-trivial instance of the GPT cryptosystem, consider the field $\mathbb{F}_{2^5}$ (where $q = 2$ and $m = 5$). Alice selects code parameters $n = 5$, $k = 1$ and $\ell = 1$, and generates a Gabidulin code $\mathcal{G}$ defined by the systematic generator matrix $G = \begin{pmatrix}1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4\end{pmatrix}$, where α is a primitive element of $\mathbb{F}_{2^5}$. This code has a minimum rank distance of $d = n - k + 1 = 5$, giving it an error-correcting capability of $t = 2$.

For the keys, Alice selects a trivial scrambler matrix $S = (1)$ and generates a distortion matrix $X \in \mathbb{F}_{2^5}^{1\times 5}$ of rank $s = 1$. Let $X = \begin{pmatrix}1 & 0 & 0 & 0 & 0\end{pmatrix}$. She computes the public key matrix $\hat{G} = S(G + X) = \begin{pmatrix}0 & \alpha & \alpha^2 & \alpha^3 & \alpha^4\end{pmatrix}$. Thus, the public key is $(\hat{G}, \hat{t} = 1)$.

Bob wishes to encrypt the message $\mathbf{x} = (1)$. He chooses a random error vector $\mathbf{e}$ with $\mathrm{rk}_q(\mathbf{e}) = \hat{t} = 1$, for example, $\mathbf{e} = (0, \alpha, 0, 0, 0)$. He then computes the ciphertext $\mathbf{c} = \mathbf{x}\hat{G} + \mathbf{e} = (0, \alpha, \alpha^2, \alpha^3, \alpha^4) + (0, \alpha, 0, 0, 0) = (0, 0, \alpha^2, \alpha^3, \alpha^4)$.

Upon reception, Alice attempts to decode $\mathbf{c}$ using the decoding algorithm for her private code $\mathcal{G}$. The ciphertext can be expanded as $\mathbf{c} = \mathbf{x}S(G + X) + \mathbf{e} = \mathbf{x}SG + (\mathbf{x}SX + \mathbf{e})$. The term in parentheses represents the effective error pattern E that the decoder must handle. Substituting the values, we find $E = \mathbf{x}SX + \mathbf{e} = (1)(1)X + \mathbf{e} = (1, 0, 0, 0, 0) + (0, \alpha, 0, 0, 0) = (1, \alpha, 0, 0, 0)$. The rank of this error vector is $\mathrm{rk}_q(E) = 2$, as 1 and α are linearly independent over the base field $\mathbb{F}_2$. Since the error-correcting capability of $\mathcal{G}$ is two, the decoding algorithm succeeds. It returns the codeword $\mathbf{c}' = \mathbf{x}SG = \mathbf{c} - E = (1, \alpha, \alpha^2, \alpha^3, \alpha^4)$. Finally, Alice recovers the message $\mathbf{x} = \mathbf{c}' \cdot S^{-1}$ (extracting the first component since G is systematic), obtaining $\mathbf{x} = (1)$. □

In Section 4.1, we will give the reason why smaller public key sizes are possible in the GPT cryptosystem as compared to McEliece. However, the security of the scheme depends heavily on the indistinguishability of the public matrix $\hat{G}$. Unlike the Hamming metric, where random codes are abundant and hard to decode, the structured nature of MRD codes makes them potentially easier to distinguish from random codes. If the structure of the underlying code $\mathcal{C}$ is not sufficiently masked by S and X, an attacker may be able to recover the private key. As we will see, the original GPT system and many of its direct successors were found to be vulnerable to such structural attacks.

3 Maximum-rank distance codes

Maximum-rank distance (MRD) codes are rank-metric codes that achieve optimal error detection and correction capabilities. For over thirty years, Gabidulin codes [28] were the only known family of such codes. However, since Sheekey proposed a new family of MRD codes in 2016 [94], multiple new constructions have emerged. Geometric connections play a crucial role in this development; they provide alternate construction methods for rank-metric codes and define the geometric constraints necessary for the constructed codes to possess the MRD property. The geometric objects can subsequently be used to determine (in)equivalence.

Research in this area is primarily driven by two applications. The first is the main motivation for this thesis: code-based cryptography. Although the usage of rank-metric codes offers the potential for smaller key sizes, Gabidulin and related codes are often too structured, making them vulnerable to structural attacks. Consequently, new and larger families of MRD codes are necessary to resist these cryptographic attacks.

The second application where (maximum) rank distance codes play a significant role is random network coding. In simple terms, a network can be modelled as a directed multigraph consisting of sources, intermediate nodes, and sinks. Source nodes transmit messages to sinks via intermediate nodes, where the mixing of different packets (e.g. by addition) occurs. This mixing is the core concept of network coding. A receiver collects the data packets and attempts to recover the messages originally intended for the sinks. First, the seminal paper of Kötter and Kschischang [50] studied error correction in such networks and introduced the framework of subspace codes to address it. Shortly thereafter, the link to rank-metric codes for constructing such subspace codes was established in [98]. Since then, this application has remained a major motivation for the study of rank-metric codes.

In this chapter, we recall the definition of MRD codes and present a few fundamental results. We then discuss some algebraic representations of these codes. A dedicated section is provided for the Gabidulin family and its generalisations. The majority of the chapter is reserved for the study of rank-metric codes through a geometric lens, presenting a connection with scattered subspaces and linear sets that advances both domains. Finally, some methods for constructing the necessary scattered subspaces are briefly considered.

3.1 Definition and first results

Although maximum-rank distance codes were introduced in the previous chapter, we restate their formal definition here, as they are the main object of study in this chapter. Recall that the codewords in a rank-metric code are linear maps, that the rank of such a map is the dimension of its image, and that this defines a valid metric.

Definition 3.1.1 (Maximum-rank distance code). Let $\mathbb{F}_q^m$ and $\mathbb{F}_q^n$ be vector spaces of dimension m and n over $\mathbb{F}_q$, respectively. A rank-metric code $\mathcal{C} \subseteq \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$ with minimum distance d for which

$$|\mathcal{C}| = q^{\max\{m,n\}(\min\{m,n\}-d+1)}$$

is called a *maximum-rank distance* (MRD) code.

Evidently, MRD codes are the rank-metric analogue of MDS codes in the Hamming metric, as both families are optimal with respect to the Singleton bound in their respective settings. However, there is a significant distinction regarding their existence. It is a well-known fact that MDS codes do not

exist for all parameter values q, n, and d. For instance, it was proven in [49] that an MDS code over $\mathbb{F}_7$ with parameters $n = 9$, $k = 7$ and $d = 3$ must be equivalent to a linear code; yet the existence of such a linear code is ruled out by [2, Corollary 9.1], a specific (proven) case of the broader (linear) MDS conjecture.[1] In contrast, MRD codes exist for all finite fields $\mathbb{F}_q$ and all parameters n, m, and d, as established by Delsarte [21].

Theorem 3.1.2 ([21, Theorem 6.3]). *There exists an MRD code in* $\mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$ *with minimum distance* d *for all* q, m, n, *and* $1 \leq d \leq \min\{m, n\}$.

In the Hamming metric, linear codes are the primary objects of study. In the rank metric, however, a more refined distinction regarding linearity is required, as already noted in Remark 2.1.9. This distinction arises from the internal structure of the finite field. Recall that $\mathbb{F}_q$ has characteristic p, so that $q = p^h$ for some prime p and integer $h \geq 1$. Consequently, $\mathbb{F}_q$ is naturally a vector space over its prime subfield $\mathbb{F}_p$. This leads to two possible notions of linearity: linearity over the base field or over a field extension. Codes that are only linear over the base (prime) field will be referred to as *additive* codes.

A crucial property of additive rank-metric codes is that their distance structure behaves identically to that of linear Hamming-metric codes.

Proposition 3.1.3. *Let* $\mathcal{C}$ *be an additive rank-metric code. The minimum rank distance of* $\mathcal{C}$ *is equal to the minimum rank weight of its non-zero codewords:*

$$d(\mathcal{C}) = \min_{c \in \mathcal{C},\, c \neq 0} \mathrm{rank}(c).$$

Proof. By definition, $d(\mathcal{C}) = \min\{\mathrm{rank}(x - y) : x, y \in \mathcal{C}, x \neq y\}$. Since $\mathcal{C}$ is additive, it is closed under subtraction; thus for any $x, y \in \mathcal{C}$, the difference $c = x - y$ belongs to $\mathcal{C}$. As $x \neq y$, we have $c \neq 0$. Therefore, the set of mutual distances corresponds exactly to the set of rank weights of non-zero codewords. ∎

Remark 3.1.4. In the following sections, we will often consider $\mathbb{F}_q$, with $q = p^h$, as the base field, and $\mathbb{F}_{q^m}$ ($m > 1$) as a field extension. In this setting, we will distinguish between $\mathbb{F}_q$-linearity, $\mathbb{F}_{q^m}$-linearity and additivity ($\mathbb{F}_p$-linearity).

A further result from [21] concerns the rank distribution, which is the rank-metric equivalent of the weight distribution in the Hamming metric.

Definition 3.1.5 (Rank distribution). For a rank-metric code $\mathcal{C} \subseteq \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$ with $m' := \min\{m, n\}$, the *rank distribution* is an $(m' + 1)$-tuple of non-negative integers $(A_0, A_1, \ldots, A_{m'})$, where

$$A_i = |\{ f \in \mathcal{C} : \mathrm{rank}(f) = i \}|.$$

The *rank enumerator* of $\mathcal{C}$, denoted by $W_{\mathcal{C}}(x, y)$, is defined as the polynomial

$$W_{\mathcal{C}}(x, y) = \sum_{i=0}^{m'} A_i x^i y^{m'-i}.$$

Interestingly, the rank distribution of an additive MRD code is completely determined by its parameters. The formula uses Gaussian coefficients.

Definition 3.1.6 (Gaussian binomial coefficient). For integers $n \geq k \geq 0$, the *Gaussian binomial coefficient* is defined as

$$\begin{bmatrix} n \\ k \end{bmatrix}_q := \frac{(q^n - 1)(q^n - q) \cdots (q^n - q^{k-1})}{(q^k - 1)(q^k - q) \cdots (q^k - q^{k-1})}.$$

If $k < 0$ or $k > n$, we set the value to zero.

[1] Let $\mathcal{C}$ be a linear $[n, k, d]$-code over the finite field $\mathbb{F}_q$ satisfying the Singleton bound (in the Hamming metric). Then $n \leq q + 1$, unless q is even and $k = 3$ or $k = q - 1$, in which case $n \leq q + 2$.

Theorem 3.1.7 ([21, Theorem 5.6]). *Let $\mathcal{C} \subseteq \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$ be an additive MRD code with minimum distance d. Let $m' = \min\{m, n\}$ and $n' = \max\{m, n\}$. Its rank distribution is then given by*

$$A_i = \begin{bmatrix} m' \\ i \end{bmatrix}_q \sum_{t=0}^{i-d} (-1)^{t+d-i} \begin{bmatrix} i \\ i-d-t \end{bmatrix}_q q^{\binom{i-d-t}{2}} (q^{n'(t+1)} - 1),$$

for $i \geq d$, with $A_0 = 1$ and $A_i = 0$ for $0 < i < d$.

The following theorem, derived from the MacWilliams identities for the rank metric, is used later.

Theorem 3.1.8 ([89, Proof of Corollary 44]). *Let $\mathcal{C}$ be an MRD code with parameters $(m, n, q; d)$. Let $m' = \min\{m, n\}$ and $n' = \max\{m, n\}$. Then for any $\nu \in \{0, \ldots, m' - d\}$, we have*

$$\begin{bmatrix} m' \\ \nu \end{bmatrix}_q + \sum_{i=d}^{m'-\nu} A_i \begin{bmatrix} m'-i \\ \nu \end{bmatrix}_q = \frac{|\mathcal{C}|}{q^{n'\nu}} \begin{bmatrix} m' \\ \nu \end{bmatrix}_q.$$

Finally, in the study of rank-metric codes, the left and right idealisers, introduced in [56], are important invariants to determine code equivalence.

Definition 3.1.9 (Code equivalence). Two rank-metric codes $\mathcal{C}, \mathcal{C}' \subseteq \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$ are said to be *equivalent* if there exist $g \in \mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_q^n)$, $h \in \mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_q^m)$, $k \in \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$, with g, h invertible, and a field automorphism ρ of $\mathbb{F}_q$ such that

$$\mathcal{C}' = \{\, g \circ f^\rho \circ h + k : f \in \mathcal{C} \,\}.$$

When $\mathcal{C}$ and $\mathcal{C}'$ are additive, they are *(linearly) equivalent* if and only if there exists an equivalence such that $k = 0$.

Remark 3.1.10. A subtle but crucial distinction from the Hamming metric arises when defining equivalence. In the Hamming metric, any distance-preserving bijection between two codes can naturally be extended to a global isometry of the entire vector space. In the rank metric, however, this is not true. It is possible to have two rank-metric codes with a rank-preserving bijection between their codewords, yet no global isometry exists that transforms one code into the other (for an explicit counterexample, see e.g. [91, Example 1.3.1]). Consequently, rank-metric code equivalence must be defined via the existence of a global transformation of the full space $\mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$.

Definition 3.1.11 (Idealiser). The *left* and *right idealisers* $L(\mathcal{C})$ and $R(\mathcal{C})$ of a rank-metric code $\mathcal{C} \subseteq \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$ are defined as[2]

$$L(\mathcal{C}) = \left\{\, \rho \in \mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_q^m) : f \circ \rho \in \mathcal{C}, \text{ for all } f \in \mathcal{C} \,\right\},$$
$$R(\mathcal{C}) = \left\{\, \rho \in \mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_q^n) : \rho \circ f \in \mathcal{C}, \text{ for all } f \in \mathcal{C} \,\right\}.$$

The following results regarding idealisers were proven in [66].

Theorem 3.1.12. *Let $\mathcal{C}$ and $\mathcal{C}'$ be $\mathbb{F}_q$-linear rank-metric codes of $\mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$.*

(i) If $\mathcal{C}$ and $\mathcal{C}'$ are equivalent, then their left and right idealisers are isomorphic as $\mathbb{F}_q$-algebras ([66, Proposition 4.1]).

(ii) Let $\mathcal{C}$ have minimum distance $d > 1$. If $m \leq n$, then $L(\mathcal{C})$ is a finite field with $|L(\mathcal{C})| \leq q^m$. If $m \geq n$, then $R(\mathcal{C})$ is a finite field with $|R(\mathcal{C})| \leq q^n$. In particular, when $m = n$, $L(\mathcal{C})$ and $R(\mathcal{C})$ are both finite fields ([66, Theorem 5.4 and Corollary 5.6]).

Invariants such as the rank distribution, equivalence relations, and idealisers are essential tools for distinguishing between MRD codes. They allow us to verify when a proposed construction is genuinely new, rather than equivalent to an existing one.

[2] Since we are working with row spaces, composition corresponds to matrix multiplication as follows: pre-composition $f \circ \rho$ corresponds to left multiplication ρf, while post-composition $\rho \circ f$ corresponds to right multiplication $f\rho$. Note that $L(\mathcal{C})$ acts on the domain and $R(\mathcal{C})$ on the codomain.

3.2 Algebraic representations of rank-metric codes

As defined previously, rank-metric codes are subsets of the space $\mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$ of $\mathbb{F}_q$-homomorphisms from $\mathbb{F}_q^m$ to $\mathbb{F}_q^n$. While this coordinate-free definition provides a nice theoretical framework, practical applications and proofs often require more concrete descriptions. In this section, we present three distinct algebraic representations of these abstract maps: matrices, vectors, and linearised polynomials. Even more representations are used throughout the literature; for an overview and how to translate between them, see [95].

Notation 3.2.1. Throughout this text, vectors $\mathbf{v} \in \mathbb{F}_q^k$ are treated as row vectors. Accordingly, linear maps are represented by matrices acting on the right (i.e. $\mathbf{v} \mapsto \mathbf{v}M$). All vector space operations are thus performed within the context of row spaces.

It is important to note that these representations are not always equivalent in scope. While the matrix representation is universal, others – such as linearised polynomials – impose additional structural requirements, for instance requiring the underlying vector spaces to possess the structure of a finite field extension (e.g. identifying $\mathbb{F}_q^n$ with $\mathbb{F}_{q^n}$). This diversity of perspectives establishes a rich interplay between domains. Not only can we leverage tools from other domains to construct and analyse codes, but problems originating in (rank-metric) coding theory frequently inspire new insights and results in these other mathematical areas.

We will restrict our discussion in this section to the three aforementioned representations. In later sections and chapters, we will refer to rank-metric codes without specifying the representation used, unless necessary.

3.2.1 Seen as matrices

The natural way to represent $\mathbb{F}_q$-homomorphisms from $\mathbb{F}_q^m$ to $\mathbb{F}_q^n$, i.e. $\mathbb{F}_q$-linear maps, is through the use of matrices. This framework is the original setting in which Delsarte initiated the study of rank-metric codes. To be precise, Delsarte used bilinear forms and their matrix representations in his seminal paper [21].

Definition 3.2.2. Let $\mathbb{F}_q^{m\times n}$ be the set of $m \times n$ matrices with entries in $\mathbb{F}_q$. The rank weight of a matrix $X \in \mathbb{F}_q^{m\times n}$ is defined to be its rank, denoted by $\mathrm{rank}(X)$. The rank distance between two matrices $X, Y \in \mathbb{F}_q^{m\times n}$ is defined as

$$d_R(X, Y) = \mathrm{rank}(X - Y).$$

A *(matrix) rank-metric code* is then a subset $\mathcal{C}$ of the metric space $(\mathbb{F}_q^{m\times n}, d_R)$. If $\mathcal{C}$ is an $\mathbb{F}_q$-subspace, we call the code linear.

Since the rank of a matrix is invariant under a change of basis of the underlying vector spaces $\mathbb{F}_q^m$ and $\mathbb{F}_q^n$, this metric is well-defined and provides a valid equivalent representation of the abstract space $\mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$ without restrictions.

Equivalence between rank-metric codes is easily expressed in the matrix representation. We say that two rank-metric codes $\mathcal{C}, \mathcal{C}' \subseteq \mathbb{F}_q^{m\times n}$ are *equivalent* if and only if there exist $X \in \mathrm{GL}_m(\mathbb{F}_q)$, $Y \in \mathrm{GL}_n(\mathbb{F}_q)$, $Z \in \mathbb{F}_q^{m\times n}$, and a field automorphism ρ of $\mathbb{F}_q$ such that

$$\mathcal{C}' = \{\, XC^\rho Y + Z : C \in \mathcal{C} \,\}.$$

When $\mathcal{C}$ and $\mathcal{C}'$ are linear, we can assume $Z = 0$, such as in Definition 3.1.9.

Furthermore, the matrix representation is the most natural setting to discuss duality of a linear rank-metric code.

Definition 3.2.3 (Delsarte dual code). The *(Delsarte) dual code* $\mathcal{C}^\perp$ of an $\mathbb{F}_q$-linear rank-metric code $\mathcal{C} \subseteq \mathbb{F}_q^{m\times n}$ is given by

$$\mathcal{C}^\perp = \left\{ N \in \mathbb{F}_q^{m\times n} : \langle M, N\rangle = 0 \text{ for each } M \in \mathcal{C} \right\},$$

where $\langle\cdot,\cdot\rangle$ denotes a symmetric bilinear form on $\mathbb{F}_q^{m\times n}$ defined by

$$\langle M, N\rangle = \mathrm{tr}(MN^\top),$$

with $N^\top$ the transpose matrix of N.

Using the theory of association schemes, Delsarte [21] proved the following result, showing that the MRD property is kept under duality.

Theorem 3.2.4 ([21, Theorem 5.5]). *Let $\mathcal{C} \subseteq \mathbb{F}_q^{m\times n}$ be an $\mathbb{F}_q$-linear MRD code of dimension k with minimum distance $d > 1$. Then the dual code $\mathcal{C}^\perp \subseteq \mathbb{F}_q^{m\times n}$ is an MRD code of dimension $nm - k$ with minimum distance* $\min\{m, n\} - d + 2$.

In general, rank-metric codes are frequently introduced directly as matrix codes. It should also be noted that this representation is favoured for its universality; unlike frameworks requiring specific algebraic constraints on dimensions, the matrix setting applies without restriction to any m and n.

3.2.2 Seen as vectors

We can equivalently view a rank-metric code in terms of vectors in the vector space $\mathbb{F}_{q^m}^n$. So, we consider n-dimensional (row) vectors over $\mathbb{F}_{q^m}$.

Definition 3.2.5. Let $\mathbb{F}_{q^m}^n$ be the set of n-dimensional vectors over $\mathbb{F}_{q^m}$. The rank weight of a vector $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{F}_{q^m}^n$ is the dimension of the vector space generated over $\mathbb{F}_q$ by its entries, i.e. $\mathrm{wt}_R(\mathbf{v}) = \dim_{\mathbb{F}_q}(\langle v_1, \ldots, v_n\rangle_{\mathbb{F}_q})$. The rank distance between two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{F}_{q^m}^n$ is defined as

$$d_R(\mathbf{v}, \mathbf{w}) = \mathrm{wt}_R(\mathbf{v} - \mathbf{w}) = \dim_{\mathbb{F}_q}(\langle v_1 - w_1, \ldots, v_n - w_n\rangle_{\mathbb{F}_q}).$$

A *(vector) rank-metric code* is then a subset $\mathcal{C}$ of the metric space $(\mathbb{F}_{q^m}^n, d_R)$. If $\mathcal{C}$ is an $\mathbb{F}_{q^m}$-subspace, we call the code linear.

Notation 3.2.6. A linear vector rank-metric code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ is called an $[n, k, d]_{q^m/q}$-code if $k = \dim_{\mathbb{F}_{q^m}}(\mathcal{C})$ and d is its minimum rank distance.

Remark 3.2.7. In this context, a linear code is traditionally defined using $\mathbb{F}_{q^m}$-linearity. In the matrix representation, it was defined using $\mathbb{F}_q$-linearity.

The extra restriction on linearity leads to the following definition of equivalence. Two $\mathbb{F}_{q^m}$-linear rank-metric codes $\mathcal{C}, \mathcal{C}' \subseteq \mathbb{F}_{q^m}^n$ are equivalent if and only if there exists $A \in \mathrm{GL}_n(\mathbb{F}_q)$ such that $\mathcal{C}' = \mathcal{C}A = \{ cA : c \in \mathcal{C} \}$.

Furthermore, usually, it is assumed that a vector rank-metric code is non-degenerate, that is, it is not artificially "inflated" by linearly dependent columns.

Definition 3.2.8 (Non-degenerate). An $[n, k, d]_{q^m/q}$-code $\mathcal{C}$ is said to be *non-degenerate* if the columns of any generator matrix of $\mathcal{C}$ are $\mathbb{F}_q$-linearly independent.

Finally, the Singleton-like bound in this setting equals, for $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ an $[n, k, d]_{q^m/q}$-code,

$$mk \leq \max\{m, n\}(\min\{m, n\} - d + 1). \tag{3.1}$$

Hence, an $[n, k, d]_{q^m/q}$-code is called a ($\mathbb{F}_{q^m}$-linear) maximum-rank distance code if its parameters attain the bound (3.1).

3.2.3 Seen as linearised polynomials

A final algebraic representation that cannot be overlooked, due to its extensive use in constructing new codes, is that of linearised polynomials.

Definition 3.2.9 (Linearised polynomial). Let $\sigma\colon x \mapsto x^{q^s}$ be a generator of the Galois group $\mathrm{Gal}(\mathbb{F}_{q^m}/\mathbb{F}_q)$, where $1 \leq s \leq m$ and $\gcd(m, s) = 1$. A polynomial of the form

$$L(x) = \sum_{i=0}^{r} \alpha_i x^{\sigma^i},$$

with coefficients $\alpha_i \in \mathbb{F}_{q^m}$ and r a non-negative integer, is called a *(linearised) σ-polynomial*. The set of all σ-polynomials over $\mathbb{F}_{q^m}$ is denoted by $\mathcal{L}_{m,\sigma}[x]$. Furthermore, if $\alpha_r \neq 0$, we say that r is the *σ-degree* of $L(x)$ and write $\deg_\sigma(L) = r$.

Notation 3.2.10. In the definition above, if $s = 1$ (i.e. $\sigma\colon x \mapsto x^q$), we use the notation $\mathcal{L}_{m,q}[x]$ and refer to these simply as linearised polynomials. The set of σ-polynomials with σ-degree less than m is denoted by

$$\widetilde{\mathcal{L}}_{m,\sigma}[x] := \left\{ \sum_{i=0}^{m-1} \alpha_i x^{\sigma^i} : \alpha_i \in \mathbb{F}_{q^m} \right\}.$$

This set is identified with the ring of σ-polynomials taken modulo $x^{\sigma^m} - x$.

Any such linearised σ-polynomial $L(x)$ satisfies $L(\beta + \gamma) = L(\beta) + L(\gamma)$ for all $\beta, \gamma \in \mathbb{F}_{q^m}$ and $L(c\beta) = cL(\beta)$ for all $c \in \mathbb{F}_q$. Thus, if $\mathbb{F}_{q^m}$ is considered as a vector space over $\mathbb{F}_q$, the linearised σ-polynomial $L(x)$ induces an $\mathbb{F}_q$-linear operator on $\mathbb{F}_{q^m}$. To be precise, there exists a one-to-one correspondence.

Proposition 3.2.11. *Let σ be a generator of* $\mathrm{Gal}(\mathbb{F}_{q^m}/\mathbb{F}_q)$*. Every $\mathbb{F}_q$-linear map from $\mathbb{F}_{q^m}$ to $\mathbb{F}_{q^m}$ can be uniquely represented as a linearised σ-polynomial of σ-degree at most $m-1$, i.e. as*

$$f\colon x \mapsto \alpha_0 x + \alpha_1 x^{\sigma} + \cdots + \alpha_{m-1} x^{\sigma^{m-1}} \in \widetilde{\mathcal{L}}_{m,\sigma}[x],$$

where $\alpha_i \in \mathbb{F}_{q^m}$ for $0 \leq i \leq m-1$.

Proof. First of all, we verify that every $L \in \widetilde{\mathcal{L}}_{m,\sigma}[x]$ induces an $\mathbb{F}_q$-linear map. Since the map $x \mapsto x^{\sigma^i}$ is an automorphism of $\mathbb{F}_{q^m}$ fixing $\mathbb{F}_q$, we have, for all $\beta, \gamma \in \mathbb{F}_{q^m}$, that

$$L(\beta+\gamma) = \sum_{i=0}^{m-1} \alpha_i (\beta+\gamma)^{\sigma^i} = \sum_{i=0}^{m-1} \alpha_i (\beta^{\sigma^i} + \gamma^{\sigma^i}) = L(\beta) + L(\gamma),$$

where we used the additivity of automorphisms. Furthermore, for all scalars $c \in \mathbb{F}_q$, it holds that $c^{\sigma^i} = c$ for $0 \leq i \leq m-1$. Thus,

$$L(c\beta) = \sum_{i=0}^{m-1} \alpha_i (c\beta)^{\sigma^i} = \sum_{i=0}^{m-1} \alpha_i c^{\sigma^i} \beta^{\sigma^i} = c \sum_{i=0}^{m-1} \alpha_i \beta^{\sigma^i} = cL(\beta).$$

We now show that this representation is unique and covers all linear maps. For this purpose, define a mapping $\Psi\colon \widetilde{\mathcal{L}}_{m,\sigma}[x] \to \mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_{q^m})$. It holds that $|\mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_{q^m})| = q^{m^2}$. The set $\widetilde{\mathcal{L}}_{m,\sigma}[x]$ is determined by the m coefficients $\alpha_0, \ldots, \alpha_{m-1}$ chosen from $\mathbb{F}_{q^m}$. Thus, the cardinality of $\widetilde{\mathcal{L}}_{m,\sigma}[x]$ equals $|\mathbb{F}_{q^m}|^m = (q^m)^m = q^{m^2}$. Because the finite sets have the same cardinality, it suffices to show that Ψ is injective. Suppose $L \in \widetilde{\mathcal{L}}_{m,\sigma}[x]$ induces the zero map, i.e. $L(\beta) = 0$ for all $\beta \in \mathbb{F}_{q^m}$. Because σ generates the Galois group, the terms x^{σ^i} correspond to the monomials x^{q^j}, $0 \leq j \leq m-1$, in some

order. Hence, viewed as a standard polynomial, $L(x)$ has algebraic degree at most q^{m-1}. However, it has q^m distinct roots (all elements of $\mathbb{F}_{q^m}$). Since a non-zero polynomial over a field cannot have more roots than its degree, $L(x)$ must be the zero polynomial. Therefore, Ψ is injective. We conclude that Ψ is a bijection, proving the claim. ∎

This means we can use linearised polynomials to represent rank-metric codes. In fact, the use of linearised polynomials for rank-metric codes was established by Gabidulin in 1985 when he (re)constructed (after Delsarte [21][3]) the first MRD code: the Gabidulin code [28]. After a long period of stagnation, Sheekey [94] demonstrated the versatility of this framework in 2016 by constructing twisted Gabidulin codes, a new family of MRD codes, using these polynomials. This breakthrough triggered a resurgence of results using linearised polynomials, including generalised twisted Gabidulin codes [65], additive constructions [75] and further distinct families [18, 60, 100]. To this day, this representation serves as an important tool for defining and analysing (new and) modern rank-metric codes.

Remark 3.2.12. Although we restrict our attention to the case $s = 1$ (i.e. $\sigma\colon x \mapsto x^q$) for the remainder of this section, it is worth noting that the theory of rank-metric codes can be developed using general σ-polynomials. While Proposition 3.2.11 implies that the set of represented endomorphisms remains the same regardless of s, the choice of σ alters the algebraic coefficients of the codewords. Allowing for $s \neq 1$ was the first generalisation of Gabidulin codes.

While often used, it must be noted that this representation only works for the case of square matrices, so when $m = n$. Proposition 3.2.11 shows us that every endomorphism $f \in \mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_{q^m})$ can be represented by a linearised polynomial with σ-degree at most $m - 1$.

Definition 3.2.13. The rank weight of a linearised polynomial $f \in \widetilde{\mathcal{L}}_{m,q}[x]$ is the rank of its corresponding endomorphism. The rank distance between two polynomials $f, g \in \widetilde{\mathcal{L}}_{m,q}[x]$ is defined as

$$d_R(f, g) = \mathrm{rank}(f - g).$$

A *(polynomial) rank-metric code* is then a subset $\mathcal{C}$ of the metric space $(\widetilde{\mathcal{L}}_{m,q}[x], d_R)$. If $\mathcal{C}$ is an $\mathbb{F}_q$-subspace, we call the code linear.

It is a known result that $\mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_{q^m})$ forms a vector space over $\mathbb{F}_{q^m}$ under the operations of addition $(f + g)(x) := f(x) + g(x)$ and scalar multiplication $(\alpha f)(x) := \alpha f(x)$. Since there are q^{m^2} such maps, this space is isomorphic to $V(m, q^m)$. Therefore, an $\mathbb{F}_{q^m}$-linear polynomial rank-metric code $\mathcal{C} \subseteq \widetilde{\mathcal{L}}_{m,q}[x]$ of dimension k is classified as an MRD code if $|\mathcal{C}| = q^{mk}$ and the rank of any non-zero $f \in \mathcal{C}$ is at least $m - k + 1$; see Proposition 3.1.3.

Equivalence can also be translated straightforwardly to the setting of linearised polynomials. We say that two polynomial rank-metric codes $\mathcal{C}, \mathcal{C}' \subseteq \widetilde{\mathcal{L}}_{m,q}[x]$ are equivalent if there exist two invertible linearised polynomials $f, g \in \mathcal{L}_{m,q}[x]$ and a field automorphism ρ of $\mathbb{F}_{q^m}$ such that

$$\mathcal{C}' = f \circ \mathcal{C}^\rho \circ g = \{\, f \circ h^\rho \circ g : h \in \mathcal{C} \,\},$$

where $h^\rho := \sum_{i=0}^{m-1} \rho(\alpha_i) x^{q^i}$ if $h(x) = \sum_{i=0}^{m-1} \alpha_i x^{q^i}$.

For $\mathbb{F}_{q^m}$-linear codes that contain an invertible element $f(x) \in \mathcal{C}$, a special form is provided.

Definition 3.2.14 (Canonical form). Let $\mathcal{C} = \langle f_1(x), \ldots, f_k(x) \rangle_{\mathbb{F}_{q^m}} \subseteq \widetilde{\mathcal{L}}_{m,q}[x]$ with $\dim_{\mathbb{F}_{q^m}}(\mathcal{C}) = k$. If there exists an invertible element $f(x) \in \mathcal{C}$, then $\mathcal{C}$ is equivalent to a rank-metric code $\mathcal{C}'$ in *canonical form*

$$\mathcal{C}' = \langle x, g_2(x), \ldots, g_k(x) \rangle_{\mathbb{F}_{q^m}},$$

for some $g_2, \ldots, g_k \in \widetilde{\mathcal{L}}_{m,q}[x]$.

This canonical form provides a convenient normalisation for certain rank-metric codes.

[3] This is an example of Stigler's law of eponymy which states that no scientific discovery is named after its original discoverer, and is a self-confirming hypothesis as Stigler's law is attributed to Robert K. Merton.

3.2.4 Translating between different representations

The three algebraic representations of rank-metric codes discussed above are mutually compatible, allowing us to translate results and properties between these frameworks (see Figure 3.1). However, a crucial distinction regarding their scope must be made: while the correspondence between the matrix and vector frameworks is universal for any parameters m and n, the standard linearised polynomial representation is restricted to the square case where $m = n$.

To establish the link between the vector space $\mathbb{F}_{q^m}^n$ and the matrix space $\mathbb{F}_q^{m\times n}$, we use the underlying field structure. Let $B = (b_1, \dots, b_m)$ be an ordered $\mathbb{F}_q$-basis for the extension field $\mathbb{F}_{q^m}/\mathbb{F}_q$. Every element $\alpha \in \mathbb{F}_{q^m}$ has a unique representation as $\alpha = \sum_{j=1}^m \alpha_j b_j$, where $\alpha_j \in \mathbb{F}_q$. We denote the coordinate vector of α with respect to B as $\nu_B(\alpha) = (\alpha_1, \dots, \alpha_m)$. We define the map $\Gamma_q \colon \mathbb{F}_{q^m}^n \to \mathbb{F}_q^{m\times n}$, which replaces each entry v_i of a vector $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{F}_{q^m}^n$ with its coordinate vector:

$$\Gamma_q(\mathbf{v}) = \begin{pmatrix} | & | & & | \\ \nu_B(v_1)^\top & \nu_B(v_2)^\top & \dots & \nu_B(v_n)^\top \\ | & | & & | \end{pmatrix}.$$

This map is an isometry preserving the rank distance, meaning a vector code translates to a matrix code with the same metric properties. The inverse map Γ_q^{-1} simply reverses the process: given a matrix, each column is interpreted as the coordinate vector of an element in $\mathbb{F}_{q^m}$ relative to B, reconstructing the vector $\mathbf{v}$.

To bridge the gap between linearised polynomials $\widetilde{\mathcal{L}}_{m,q}[x]$ and the vector space $\mathbb{F}_{q^m}^n$, we must work within the specific setting of $m = n$. In this context, we use an ordered basis $\mathcal{B} = (\beta_1, \dots, \beta_m)$ of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ to define the evaluation map $\mathrm{ev}_\mathcal{B} \colon \widetilde{\mathcal{L}}_{m,q}[x] \to \mathbb{F}_{q^m}^m$, given by $\mathrm{ev}_\mathcal{B}(f) = (f(\beta_1), \dots, f(\beta_m))$. This map is an isomorphism that also preserves the rank metric.

The inverse map $\mathrm{ev}_\mathcal{B}^{-1}$ assigns to a vector $\mathbf{w} = (w_1, \dots, w_m) \in \mathbb{F}_{q^m}^m$ the unique linearised polynomial $f(x) = \sum_{i=0}^{m-1} a_i x^{q^i}$ such that $f(\beta_j) = w_j$, for all $j \in \{1, \dots, m\}$. Finding this f is equivalent to solving a system of m linear equations in the unknowns $a_0, \dots, a_{m-1}$. Because the basis elements β_j are $\mathbb{F}_q$-linearly independent, this system always possesses a unique solution, ensuring that every vector in $\mathbb{F}_{q^m}^m$ corresponds to exactly one linearised polynomial.

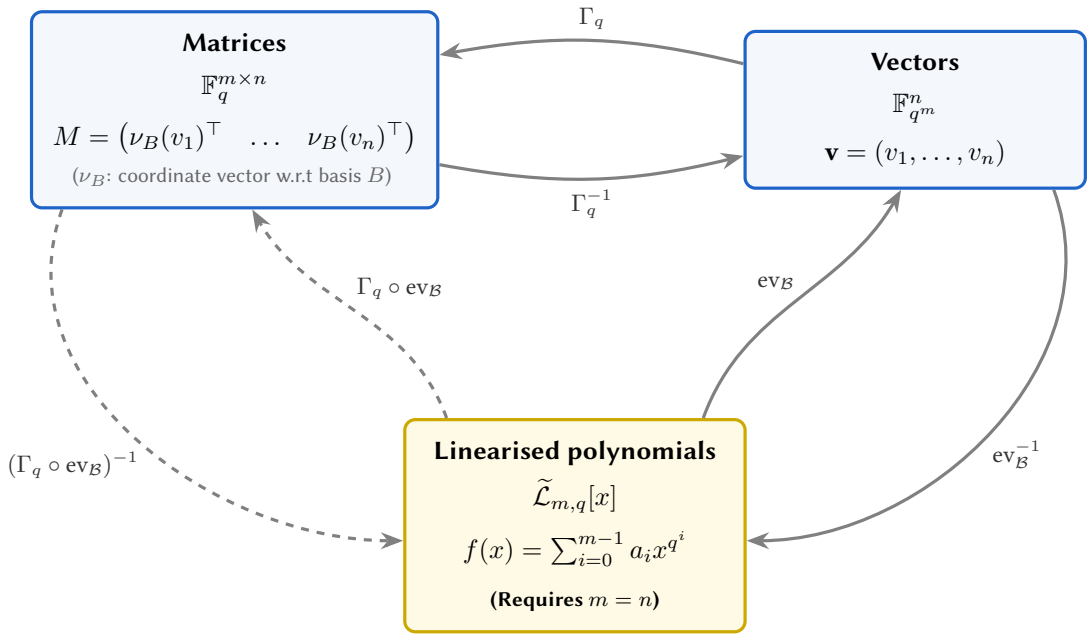


Figure 3.1: Translations between the three discussed algebraic representations of rank-metric codes.

Example 3.2.15. To illustrate these translations, consider the case $q = 2$ and $m = n = 2$. Let $\mathbb{F}_4 = \mathbb{F}_2(\alpha)$, where $\alpha^2 + \alpha + 1 = 0$, and use the basis $B = \mathcal{B} = (1, \alpha)$. Consider a code $\mathcal{C} \subseteq \mathbb{F}_2^{2\times 2}$ spanned by the matrices

$$M_1 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad M_2 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

The code $\mathcal{C} = \{0, M_1, M_2, M_1 + M_2\}$ consists of four elements. One can verify that $\det(M_1) = -1$, $\det(M_2) = -1$, and $\det(M_1 + M_2) = 1$. Since every non-zero element has full rank, the minimum distance is $d = 2$. Hence, because $|\mathcal{C}| = 2^{2(2-2+1)} = 4$, the code $\mathcal{C}$ is MRD.

Using Γ_q^{-1} with basis B, the matrix M_1 corresponds to the vector $\mathbf{v}^{(1)} \in \mathbb{F}_4^2$ where the first entry has coordinates $(0, 1)$, which is $0 \cdot 1 + 1 \cdot \alpha = \alpha$, and the second entry has coordinates $(1, 1)$, which is $1 \cdot 1 + 1 \cdot \alpha = 1 + \alpha = \alpha^2$. Thus, $\mathbf{v}^{(1)} = (\alpha, \alpha^2)$. Similarly, M_2 corresponds to the vector $\mathbf{v}^{(2)} = (1 + \alpha, 1) = (\alpha^2, 1)$.

To find the polynomial representation for M_1, we seek $f_1(x) = a_0 x + a_1 x^2$ such that $\mathrm{ev}_{\mathcal{B}}(f_1) = (\alpha, \alpha^2)$. This leads to the following system:

$$\begin{cases} a_0(1) + a_1(1)^2 = \alpha \\ a_0(\alpha) + a_1(\alpha)^2 = \alpha^2 \end{cases} \implies \begin{cases} a_0 + a_1 = \alpha \\ \alpha a_0 + \alpha^2 a_1 = \alpha^2 \end{cases}.$$

Substituting $a_1 = \alpha - a_0$ into the second equation gives $\alpha a_0 + \alpha^2(\alpha - a_0) = \alpha^2$. Recalling $\alpha^3 = 1$ and $\alpha^2 = \alpha + 1$ in $\mathbb{F}_4$, we solve to find $a_0 = \alpha$ and $a_1 = 0$. Therefore, the polynomial representation is $f_1(x) = \alpha x$. Likewise, the polynomial representation for M_2 is given by $f_2(x) = (\alpha + 1)x$. □

This example demonstrates the transition between the three frameworks. While matrices offer the most straightforward representation, and the vector representation connects to the underlying field extension, the linearised polynomial representation provides a useful algebraic toolset which we will re-encounter in Section 3.5.

3.3 Some families of MRD codes

Having established the theoretical existence of MRD codes and the various means to represent them, we now turn our attention to explicit constructions. The construction of MRD codes is a central problem in the theory of rank-metric codes. For a considerable time, the landscape of these codes was dominated by a single construction. However, recent years have witnessed a surge in the discovery of new families, largely driven by the usage of linearised polynomials and the correspondences with geometric objects.

In this section, we provide an overview of the development from classical Gabidulin codes to the modern additive generalised twisted Gabidulin codes. We place particular emphasis on the algebraic properties of these codes, specifically the distinction between $\mathbb{F}_{q^m}$-linear codes, $\mathbb{F}_q$-linear codes and those that are merely additive.

3.3.1 Gabidulin codes

The first and most fundamental class of MRD codes was introduced by Delsarte in 1978 [21] and, using linearised polynomials, by Gabidulin in 1985 [28]. These codes are the rank-metric analogue of Reed-Solomon codes in the Hamming metric. Just as Reed-Solomon codes are defined by evaluating polynomials of restricted degree, Gabidulin codes are defined by linearised polynomials of restricted σ-degree.

We work in the ambient space $\widetilde{\mathcal{L}}_{m,\sigma}[x]$, that is, the set of all linearised σ-polynomials over $\mathbb{F}_{q^m}$ with σ-degree less than m. Recall that this space is isomorphic to the ring of endomorphisms $\mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_{q^m})$ (see Proposition 3.2.11). Below is the definition of the generalised Gabidulin codes, as published in 2005 [51]. Note that the "classical" definition (by Gabidulin in 1985) corresponds to the case $s = 1$.

Definition 3.3.1 (Generalised Gabidulin code). Let k, m be integers with $1 \leq k \leq m$. Let $\sigma\colon x \mapsto x^{q^s}$ be a generator of $\mathrm{Gal}(\mathbb{F}_{q^m}/\mathbb{F}_q)$, implying $\gcd(s, m) = 1$. The *generalised Gabidulin code* $\mathcal{G}_{k,\sigma}$ is a subset of $\widetilde{\mathcal{L}}_{m,\sigma}[x]$ consisting of all σ-polynomials with σ-degree strictly less than k:

$$\mathcal{G}_{k,\sigma} = \left\{ f(x) = \sum_{i=0}^{k-1} f_i x^{\sigma^i} : f_i \in \mathbb{F}_{q^m} \right\}.$$

The algebraic structure of $\mathcal{G}_{k,\sigma}$ is straightforward. Since the sum of two polynomials with σ-degree less than k also has σ-degree less than k, and scalar multiplication by an element $\alpha \in \mathbb{F}_{q^m}$ preserves the σ-degree (as seen in the proof of Proposition 3.2.11), the generalised Gabidulin code $\mathcal{G}_{k,\sigma}$ forms a subspace over the extension field $\mathbb{F}_{q^m}$.

Proposition 3.3.2. *The generalised Gabidulin code $\mathcal{G}_{k,\sigma}$ is an $\mathbb{F}_{q^m}$-linear code of dimension k.*

Corollary 3.3.3. *Let k, m be integers with $1 \leq k \leq m$. Let $\sigma\colon x \mapsto x^{q^s}$ be a generator of $\mathrm{Gal}(\mathbb{F}_{q^m}/\mathbb{F}_q)$, implying $\gcd(s, m) = 1$. The generalised Gabidulin code $\mathcal{G}_{k,\sigma}$ can be written as*

$$\mathcal{G}_{k,\sigma} = \langle x, x^{\sigma}, \ldots, x^{\sigma^{k-1}} \rangle_{\mathbb{F}_{q^m}} \subseteq \widetilde{\mathcal{L}}_{m,\sigma}[x].$$

The most important property of $\mathcal{G}_{k,\sigma}$, however, is its minimum distance. We will prove that a generalised Gabidulin code is a maximum-rank distance code, but first we need the following lemma.

Lemma 3.3.4. *Let $f(x)$ be a non-zero σ-polynomial over $\mathbb{F}_{q^m}$ with σ-degree $\deg_\sigma(f) = r$. Then the set of roots of $f(x)$ in $\mathbb{F}_{q^m}$ forms an $\mathbb{F}_q$-linear subspace of dimension at most r. Consequently, $f(x)$ has at most q^r roots in $\mathbb{F}_{q^m}$.*

Proof. See, for example, the discussion in [55, Pages 110–111]. ∎

Theorem 3.3.5. *The generalised Gabidulin code $\mathcal{G}_{k,\sigma}$ is an MRD code. In other words, it has minimum rank distance $d = m - k + 1$.*

Proof. It is clear that $|\mathcal{G}_{k,\sigma}| = q^{mk}$ because there are, by definition, k coefficients that can take q^m values.

Now, consider any non-zero codeword $f \in \mathcal{G}_{k,\sigma}$. By definition, the σ-degree of f is at most $k - 1$. Let $V = V(m, q) \cong \mathbb{F}_{q^m}$ and consider f as a linear map on V. The kernel $\ker(f)$ is an $\mathbb{F}_q$-subspace of V. Because the number of roots of f is bounded by $q^{\deg_\sigma(f)}$, see Lemma 3.3.4, we have

$$|\ker(f)| \leq q^{k-1} \implies \dim_{\mathbb{F}_q}(\ker(f)) \leq k - 1.$$

The rank-nullity theorem states that $\mathrm{rank}(f) + \dim(\ker(f)) = m$. Therefore,

$$\mathrm{rank}(f) = m - \dim(\ker(f)) \geq m - (k - 1) = m - k + 1.$$

Using $|\mathcal{G}_{k,\sigma}| = q^{mk}$, we conclude that $\mathrm{rank}(f) = m - k + 1$ and thus $\mathcal{G}_{k,\sigma}$ is MRD. ∎

Remark 3.3.6. While the definition above focuses on the space $\widetilde{\mathcal{L}}_{m,\sigma}[x]$ (corresponding to square matrices of size $m \times m$), the family of Gabidulin codes exists for all dimensions m and n. For the general rectangular case (assuming $m \leq n$), the code is constructed by evaluating linearised polynomials over the larger field $\mathbb{F}_{q^n}$ on a set of m linearly independent elements $\alpha_1, \ldots, \alpha_m \in \mathbb{F}_{q^n}$ over $\mathbb{F}_q$. Explicitly, we obtain the (matrix) code:

$$\left\{ \begin{pmatrix} \nu_B(f(\alpha_1)) \\ \vdots \\ \nu_B(f(\alpha_m)) \end{pmatrix} : f \in \mathcal{G}_{k,\sigma} \right\}, \tag{3.2}$$

where $\nu_B(\beta) \in \mathbb{F}_q^n$ denotes the coordinate row vector of $\beta \in \mathbb{F}_{q^n}$ with respect to a fixed $\mathbb{F}_q$-basis B. Therefore, a Gabidulin code exists for any valid set of parameters m, n, q, and minimum distance d. The MRD property is retained with this construction.

Gabidulin codes possess efficient decoding algorithms. Approaches similar to the Berlekamp-Massey or Welch-Berlekamp algorithm for Reed-Solomon codes have been successfully adapted to the rank metric setting; see, for example, the work by Loidreau [57] on a Berlekamp-Welch-like algorithm. This combination of optimal parameters and practical decoding capabilities makes Gabidulin codes highly attractive for applications in random network coding and cryptography.

For nearly thirty years, Gabidulin codes (and their generalised versions involving automorphisms other than the standard Frobenius automorphism) were essentially the only known examples of MRD codes. After more than three decades, a new family of MRD codes was finally found.

3.3.2 Twisted Gabidulin codes

In 2016, Sheekey [94] provided a breakthrough in the construction of MRD codes by introducing a new family that is not equivalent to the Gabidulin family. These are known as twisted Gabidulin codes and were not much later further generalised in [65].

The construction modifies the Gabidulin structure by adding a term with a higher σ-degree, "twisted" by a coefficient dependent on the lower degree terms. The definition requires the norm of a finite field.

Definition 3.3.7 (Norm). For $\alpha \in \mathbb{F}_{q^m}$, the *norm* $\mathrm{N}_{q^m/q}(\alpha)$ of α over $\mathbb{F}_q$ is defined by

$$\mathrm{N}_{q^m/q}(\alpha) = \alpha \cdot \alpha^q \cdot \cdots \cdot \alpha^{q^{m-1}} = \alpha^{(q^m-1)/(q-1)}.$$

Definition 3.3.8 (Generalised twisted Gabidulin code). Let k, m, h be integers such that $1 \leq k, h \leq m-1$. Let $\eta \in \mathbb{F}_{q^m}$ be a fixed element such that $\mathrm{N}_{q^m/q}(\eta) \neq (-1)^{mk}$. The *generalised twisted Gabidulin code* $\mathcal{H}_{k,\sigma}(\eta, h)$ is defined as the following set of linearised polynomials in $\widetilde{\mathcal{L}}_{m,\sigma}[x]$:

$$\mathcal{H}_{k,\sigma}(\eta, h) = \left\{ f(x) = \sum_{i=0}^{k-1} f_i x^{\sigma^i} + \eta f_0^{\sigma^h} x^{\sigma^k} : f_i \in \mathbb{F}_{q^m} \right\}.$$

Note that the coefficient of the term x^{σ^k} is not free, but is determined by f_0 via the expression $\eta f_0^{\sigma^h}$. This dependency breaks $\mathbb{F}_{q^m}$-linearity: if we substitute f_0 with αf_0 (where $\alpha \in \mathbb{F}_{q^m}$), the coefficient of the twist term becomes $\eta(\alpha f_0)^{\sigma^h} = \alpha^{\sigma^h}(\eta f_0^{\sigma^h})$. This result differs from the standard scalar multiplication $\alpha(\eta f_0^{\sigma^h})$ unless $\alpha^{\sigma^h} = \alpha$. Consequently, the set is not a vector space over $\mathbb{F}_{q^m}$. However, the code remains $\mathbb{F}_q$-linear because the Frobenius automorphism acts as the identity on the base field $\mathbb{F}_q$. Thus, $\mathcal{H}_{k,\sigma}(\eta, h)$ forms a vector space over $\mathbb{F}_q$.

Remark 3.3.9. Setting $\eta = 0$ reduces the generalised twisted Gabidulin code $\mathcal{H}_{k,\sigma}(0, h)$ to the standard Gabidulin code $\mathcal{G}_{k,\sigma}$.

Sheekey [94] proved that while this construction always yields a rank-metric code, it is MRD if and only if η satisfies the norm condition (a result established for the standard Frobenius automorphism, with the general case formalised in [65]).

Lemma 3.3.10. *Let $L(x) = \sum_{i=0}^{k} a_i x^{\sigma^i} \in \mathcal{L}_{m,\sigma}[x]$ be a linearised polynomial with $\deg_\sigma(L) = k$. If $\dim_{\mathbb{F}_q}(\ker L) = k$, then the coefficients satisfy the norm condition:*

$$\mathrm{N}_{q^m/q}(a_k) = (-1)^{mk}\,\mathrm{N}_{q^m/q}(a_0).$$

Proof. See [38, Theorem 10]. ∎

Theorem 3.3.11. *The generalised twisted Gabidulin code $\mathcal{H}_{k,\sigma}(\eta, h)$ is an MRD code of size q^{mk}, that is $\dim_{\mathbb{F}_q}(\mathcal{H}_{k,\sigma}(\eta, h)) = mk$, if*

$$\mathrm{N}_{q^m/q}(\eta) \neq (-1)^{mk}.$$

Proof. Because the k coefficients $f_0, \ldots, f_{k-1}$ of a codeword in $\mathcal{H}_{k,\sigma}(\eta, h)$ can be chosen freely from $\mathbb{F}_{q^m}$, it holds that $|\mathcal{H}_{k,\sigma}(\eta, h)| = q^{mk}$.

Since $\mathcal{H}_{k,\sigma}(\eta, h)$ is an additive code, the minimum distance is determined by the minimum rank of its non-zero codewords (see Proposition 3.1.3). Let $f(x) \in \mathcal{H}_{k,\sigma}(\eta, h)$ be a non-zero codeword:

$$f(x) = \sum_{i=0}^{k-1} f_i x^{\sigma^i} + \eta f_0^{\sigma^h} x^{\sigma^k}.$$

If $f_0 = 0$, then $f(x)$ is a standard Gabidulin codeword of σ-degree at most $k-1$. Thus, it has at most q^{k-1} roots (see Lemma 3.3.4), implying $\text{rank}(f) \geq m - (k-1) = m - k + 1$, which satisfies the MRD bound.

If $f_0 \neq 0$, the polynomial has σ-degree exactly k. For the code to fail the MRD property, there must exist a codeword with $\text{rank}(f) < m - k + 1$. Because the degree is k, this is only possible if the kernel dimension is exactly k (implying $f(x)$ has q^k roots in $\mathbb{F}_{q^m}$).

By Lemma 3.3.10, kernel dimension k implies that the coefficients satisfy

$$\mathrm{N}_{q^m/q}(\eta f_0^{\sigma^h}) = (-1)^{mk} \mathrm{N}_{q^m/q}(f_0).$$

Using the multiplicativity of the norm and the fact that $\mathrm{N}_{q^m/q}(z^{\sigma^h}) = \mathrm{N}_{q^m/q}(z)$, this simplifies to

$$\mathrm{N}_{q^m/q}(\eta)\, \mathrm{N}_{q^m/q}(f_0) = (-1)^{mk} \mathrm{N}_{q^m/q}(f_0).$$

Since $f_0 \neq 0$, we can divide by $\mathrm{N}_{q^m/q}(f_0)$, yielding $\mathrm{N}_{q^m/q}(\eta) = (-1)^{mk}$. Therefore, a codeword with rank strictly less than $m - k + 1$ exists only if $\mathrm{N}_{q^m/q}(\eta) = (-1)^{mk}$. Conversely, if $\mathrm{N}_{q^m/q}(\eta) \neq (-1)^{mk}$, no such codeword exists, and the code is MRD. ∎

In [94], inequivalence was initially established for the standard Frobenius automorphism $F\colon x \mapsto x^q$, proving that the twisted code $\mathcal{H}_{k,F}(\eta, h)$ is not equivalent to the standard Gabidulin code $\mathcal{G}_{k,F}$ for all admissible parameters. This result was later generalised by Lunardon, Trombetti and Zhou [65] to arbitrary generators σ, confirming that the inequivalence holds for the broader class $\mathcal{H}_{k,\sigma}(\eta, h)$.

Intuitively, this inequivalence is demonstrated by comparing the automorphism groups of the two structures. Sheekey proved that for $k \notin \{1, m-1\}$, $h \notin \{0, 1\}$, and non-zero η, the automorphism group of the twisted code is strictly smaller than that of the standard Gabidulin code. The proof leverages the fact that the twisted code $\mathcal{H}_{k,F}(\eta, h)$ contains a unique subspace equivalent to the lower-order Gabidulin code $\mathcal{G}_{k-1,F}$. This uniqueness limits the automorphism group of the twisted code. In contrast, the standard Gabidulin code lacks this constraint and possesses a strictly larger automorphism group; this discrepancy makes the existence of an equivalence map between them impossible. Moreover, the construction of the twisted Gabidulin codes shifted the research focus from $\mathbb{F}_{q^m}$-linear subspaces to the broader, more flexible class of $\mathbb{F}_q$-linear codes.

Regarding decoding, although the structure is more complex than Gabidulin codes, efficient decoding algorithms for twisted Gabidulin codes have been developed, notably by Randrianarisoa [86], ensuring their practical relevance.

3.3.3 Additive generalised twisted Gabidulin codes

Following Sheekey's breakthrough, the mechanism of "twisting" was generalised further. It is worth noting that Otal and Özbudak [76] had independently discovered specific instances of these twisted codes (specifically the case $k = 2$) around the same time. Building on these simultaneous discoveries, Otal and Özbudak [75] provided the most comprehensive generalisation in the additive ($\mathbb{F}_{q_0}$-linear for a subfield $\mathbb{F}_{q_0}$ of $\mathbb{F}_q$) setting.

They introduced the class of additive generalised twisted Gabidulin (AGTG) codes. The construction relies on modifying the first coefficient of the generalised Gabidulin polynomial with an additive twist, subject to specific norm conditions that ensure the minimum rank property is preserved.

Definition 3.3.12 (AGTG code). Let m, k, s, h be positive integers satisfying $\gcd(s, m) = 1$ and $k < m$. Let $q = q_0^u$, $\eta \in \mathbb{F}_{q^m}$ such that $\mathrm{N}_{q^{sm}/q_0^s}(\eta) \neq (-1)^{mku}$. Then, an *additive generalised twisted Gabidulin code* is a subset of $\widetilde{\mathcal{L}}_{m,\sigma}[x]$ of the form

$$\mathcal{H}_{k,s,q_0}(\eta, h) = \left\{ \sum_{i=0}^{k-1} f_i x^{q^{si}} + \eta f_0^{q_0^h} x^{q^{sk}} : f_i \in \mathbb{F}_{q^m} \right\}.$$

Proposition 3.3.13 ([75, Theorem 3.2]). *An AGTG code $\mathcal{H}_{k,s,q_0}(\eta, h)$ is an $\mathbb{F}_{q_0}$-linear MRD code of size q^{mk} with minimum rank distance $m - k + 1$.*

Despite their complex structure, AGTG codes retain sufficient algebraic properties to allow for efficient decoding, as shown by Kadir and Li [45]. However, the loss of linearity makes it significantly more difficult to classify these codes up to equivalence. To address the classification challenges, recent research has turned to a more geometric approach involving scattered subspaces. This perspective not only facilitates the characterisation of equivalence but also enables the construction of MRD codes with parameters not achievable by polynomial-based constructions.

3.4 Rank-metric codes through a geometric lens

The families of MRD codes considered in Section 3.3 were all constructed using linearised polynomials. In particular, they are all square codes, i.e. subsets of $\mathrm{End}_{\mathbb{F}_q}(\mathbb{F}_{q^m})$. Recent work of Zini and Zullo [104] provides a different construction method using (scattered) subspaces. This allows for non-square codes and provides another way of constructing MRD codes inequivalent to Gabidulin codes (or its related families). We will follow their article closely in this section, while modifying the results to accommodate row vectors (see Notation 3.2.1) and the rank-metric code parameters defined in Definition 2.1.7.

3.4.1 Equivalence between subspaces and rank-metric codes

Throughout this section, the following notation will be used:

- $\omega_\alpha \colon \mathbb{F}_{q^m} \to \mathbb{F}_{q^m}, x \mapsto \alpha x$, for any $\alpha \in \mathbb{F}_{q^m}$;
- $\mathcal{F}_m = \{ \omega_\alpha : \alpha \in \mathbb{F}_{q^m} \}$, which is a field isomorphic to $\mathbb{F}_{q^m}$;
- $\mathcal{F}_{m,q} = \{ \omega_\alpha : \alpha \in \mathbb{F}_q \}$, which is a subfield of $\mathcal{F}_m$ isomorphic to $\mathbb{F}_q$;
- $\tau_{\mathbf{v}} \colon \mathbb{F}_{q^m} \to V, \lambda \mapsto \lambda \mathbf{v}$, for any $\mathbf{v} \in V = V(m, q)$.

It is possible to associate an $\mathbb{F}_q$-linear rank-metric code with an $\mathbb{F}_q$-subspace U of a vector space V over $\mathbb{F}_{q^m}$, as is shown in the following theorem.

Theorem 3.4.1 ([104, Theorem 3.1]). *Let $V = V(r, q^m)$ and $W = V(rm - k, q)$. Let $U = V(k, q)$ be an $\mathbb{F}_q$-subspace of V, and $G \colon V \to W$ be an $\mathbb{F}_q$-linear map with $\ker(G) = U$. Define, for any $\mathbf{v} \in V$, the $\mathbb{F}_q$-linear map*

$$\Gamma_{\mathbf{v}} = G \circ \tau_{\mathbf{v}}. \tag{3.3}$$

Furthermore, define

$$\iota = \max\{\dim_{\mathbb{F}_q}(U \cap \langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}}) : \mathbf{v} \in V \setminus \{\mathbf{0}\}\}. \tag{3.4}$$

If $\iota < m$, then the pair (U, G) defines an $\mathbb{F}_q$-linear rank-metric code

$$\mathcal{C}_{U,G} = \{ \Gamma_{\mathbf{v}} : \mathbf{v} \in V \} \tag{3.5}$$

of dimension rm with parameters $(m, rm - k, q; m - \iota)$, whose left idealiser contains $\mathcal{F}_m$.

Proof. First of all, note that $\mathcal{C}_{U,G}$ is an $\mathbb{F}_q$-vector space. Indeed, we have $\Gamma_{\mathbf{v}}+\Gamma_{\mathbf{w}} = \Gamma_{\mathbf{v}+\mathbf{w}}$ and $\alpha\Gamma_{\mathbf{v}} = \Gamma_{\alpha\mathbf{v}}$, for any $\mathbf{v}, \mathbf{w} \in V$ and $\alpha \in \mathbb{F}_q$, because G is an $\mathbb{F}_q$-linear map by assumption.

For any $\mathbf{v} \in V$, denote $R_{\mathbf{v}} = \{ \lambda \in \mathbb{F}_{q^m} : \lambda\mathbf{v} \in U \}$. By definition, $\ker(\Gamma_{\mathbf{v}}) = R_{\mathbf{v}}$ and, when $\mathbf{v} \neq 0$, $\dim_{\mathbb{F}_q}(R_{\mathbf{v}}) = \dim_{\mathbb{F}_q}(U \cap \langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}})$. Then $\dim_{\mathbb{F}_q}(\ker(\Gamma_{\mathbf{v}})) \leq \iota$ and there exists a vector $\mathbf{u} \in V \setminus \{\mathbf{0}\}$ such that equality is reached. So, by the rank-nullity theorem and Proposition 3.1.3, the minimum distance of $\mathcal{C}_{U,G}$ is $m - \iota$.

Now, for $\mathbf{v}, \mathbf{w} \in V$, suppose $\Gamma_{\mathbf{v}} = \Gamma_{\mathbf{w}}$. Linearity implies that $\Gamma_{\mathbf{v}-\mathbf{w}}$ is the zero map, so

$$G(\lambda(\mathbf{v} - \mathbf{w})) = \mathbf{0} \quad \text{for all } \lambda \in \mathbb{F}_{q^m}.$$

If $\mathbf{v} \neq \mathbf{w}$, then $\dim_{\mathbb{F}_q}(\langle \mathbf{v} - \mathbf{w} \rangle_{\mathbb{F}_{q^m}}) = m$, which leads to

$$\dim_{\mathbb{F}_q}(U \cap \langle \mathbf{v} - \mathbf{w} \rangle_{\mathbb{F}_{q^m}}) = m.$$

This contradicts the hypothesis that $\iota < m$. So, we must have $\mathbf{v} = \mathbf{w}$. Therefore, the map $\Phi \colon V \to \mathcal{C}_{U,G}$ is injective and surjective (by definition), so $\dim_{\mathbb{F}_q}(\mathcal{C}_{U,G}) = \dim_{\mathbb{F}_q}(V) = rm$.

Finally, for any $\alpha \in \mathbb{F}_{q^m}$ and $\mathbf{v} \in V$, we have $\Gamma_{\mathbf{v}} \circ \omega_\alpha = \Gamma_{\alpha\mathbf{v}}$. Consequently, the left idealiser $L(\mathcal{C}_{U,G})$ contains $\mathcal{F}_m$. ∎

Interestingly, it is possible to characterise the codes $\mathcal{C}_{U,G}$ which are maximum-rank distance codes. To achieve this goal, the following lemma is necessary and will also be of great help later.

Lemma 3.4.2 ([66, Lemma 2.1]). *Let $\mathcal{C} \subseteq \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_{q^m}, \mathbb{F}_{q^n})$ be an additive MRD code with minimum distance d. Let $m' = \min\{m, n\}$ and $n' = \max\{m, n\}$. Then, for any $0 \leq \ell \leq m' - d$, we have $A_{d+\ell} > 0$, i.e. there exists at least one $\mathbb{F}_q$-linear map $f \in \mathcal{C}$ such that $\mathrm{rank}(f) = d + \ell$.*

Proof. According to Theorem 3.1.7, the rank distribution $(A_0, \ldots, A_{m'})$ of an additive MRD code depends solely on the parameters m, n, q and the minimum distance d. Therefore, to prove that $A_{d+\ell} > 0$ for all $0 \leq \ell \leq m' - d$, it suffices to show this property for a specific family of MRD codes with these parameters. Following the construction in Equation (3.2), let $\mathcal{C}_k$ denote the generalised Gabidulin code with $k = m - d + 1$. We consider this family because, as established in Remark 3.3.6, these codes exist for all parameters.

Without loss of generality, assume $m \leq n$, so $m' = m$. We proceed by induction on the parameter k. For $k = 1$, the code $\mathcal{C}_1$ has minimum distance $m - 1 + 1 = m$. Thus, every non-zero codeword has rank m. Since the condition $0 \leq \ell \leq m - d$ becomes $0 \leq \ell \leq 0$, we only need to show the existence of rank $d = m$, which we did.

Now, assume the statement holds for k, i.e. the code $\mathcal{C}_k$ contains codewords of all ranks $r \in \{m - k + 1, \ldots, m\}$. Consider the code $\mathcal{C}_{k+1}$. By the definition of the underlying polynomial spaces, we have the inclusion

$$\mathcal{C}_k \subset \mathcal{C}_{k+1}.$$

Therefore, $\mathcal{C}_{k+1}$ contains codewords of all ranks $\{m - k + 1, \ldots, m\}$. As $\mathcal{C}_{k+1}$ is an MRD code with minimum distance $d' = m - (k + 1) + 1 = m - k$, there must exist a codeword in $\mathcal{C}_{k+1}$ with rank exactly $m - k$.

So, $\mathcal{C}_{k+1}$ contains codewords of ranks $\{m - k\} \cup \{m - k + 1, \ldots, m\}$, which covers the full range required. By induction, the property holds for all k, and thus for all additive MRD codes. ∎

Theorem 3.4.3 ([104, Theorem 3.2]). *Let $V = V(r, q^m)$ and $W = V(rm - k, q)$. Let $U = V(k, q)$ be an $\mathbb{F}_q$-subspace of V, $G \colon V \to W$ be an $\mathbb{F}_q$-linear map with $\ker(G) = U$, $\iota = \max\{\dim_{\mathbb{F}_q}(U \cap \langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}}) : \mathbf{v} \in V \setminus \{\mathbf{0}\}\} < m$ and $\mathcal{C}_{U,G} = \{ \Gamma_{\mathbf{v}} : \mathbf{v} \in V \}$. Then $\mathcal{C}_{U,G}$ is an $\mathbb{F}_q$-linear MRD code if and only if*

$$(\iota + 1) \mid rm \quad \text{and} \quad k = \frac{\iota rm}{\iota + 1} \leq (r - 1)m.$$

In this case,

- *the parameters of* $\mathcal{C}_{U,G}$ *are* $(m, \frac{rm}{\iota+1}, q; m-\iota)$*;*
- *the left idealiser of* $\mathcal{C}_{U,G}$ *is* $\mathcal{F}_m$*;*
- *the rank distribution of* $\mathcal{C}_{U,G}$ *is*

$$A_{m-s} = \begin{bmatrix} m \\ s \end{bmatrix}_q \sum_{j=0}^{\iota-s} (-1)^j \begin{bmatrix} m-s \\ j \end{bmatrix}_q q^{\binom{j}{2}} \left(q^{\frac{rm(\iota-s-j+1)}{\iota+1}} - 1 \right),$$

for $s \in \{0, 1, \ldots, \iota\}$.

Proof. If $\dim_{\mathbb{F}_q}(U) = k > (r-1)m$, then every $\mathbf{v} \in V \setminus \{\mathbf{0}\}$ intersects U, i.e. $\dim_{\mathbb{F}_q}(U \cap \langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}}) \geq 1$, and hence $\dim_{\mathbb{F}_q}(\ker(\Gamma_{\mathbf{v}})) = \dim_{\mathbb{F}_q}(U \cap \langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}}) \geq 1$, so that $\mathcal{C}_{U,G}$ has no elements of rank m. Indeed, by applying the rank-nullity theorem to $\Gamma_{\mathbf{v}} \colon \mathbb{F}_{q^m} \to W$, we obtain $\operatorname{rank}(\Gamma_{\mathbf{v}}) = \dim_{\mathbb{F}_q}(\mathbb{F}_{q^m}) - \dim_{\mathbb{F}_q}(\ker(\Gamma_{\mathbf{v}})) < m$ in that case. Thus, by Lemma 3.4.2, $\mathcal{C}_{U,G}$ is not an MRD code.

Now, suppose $k \leq (r-1)m$. Then $rm - k \geq m$ and the Singleton-like bound of (2.2) applied to $\mathcal{C}_{U,G}$ reads

$$rm \leq (rm-k)(m-(m-\iota)+1).$$

Therefore, $\mathcal{C}_{U,G}$ is an MRD code if and only if $\iota+1$ divides rm and we obtain equality, that is $k = \frac{\iota rm}{\iota+1}$.

In this case, the parameters of $\mathcal{C}_{U,G}$ are provided by Theorem 3.4.1 and the rank distribution follows from Theorem 3.1.7. By Theorem 3.4.1, the left idealiser of $\mathcal{C}_{U,G}$ contains $\mathcal{F}_m$, so $|L(\mathcal{C}_{U,G})| \geq q^m$. Since Theorem 3.1.12 implies $|L(\mathcal{C}_{U,G})| \leq q^m$, the idealiser is equal to $\mathcal{F}_m$. ∎

The choice of the map G is unimportant for the code construction. A code $\mathcal{C}_{U,G}$ is uniquely determined by U, up to equivalence.

Proposition 3.4.4 ([104, Proposition 3.3]). *Let* $V = V(r, q^m)$ *and* $W = V(rm-k, q)$*. Let* $U = V(k, q)$ *be an* $\mathbb{F}_q$*-subspace of* V*,* $G \colon V \to W$ *and* $\overline{G} \colon V \to W$ *be two* $\mathbb{F}_q$*-linear maps with* $\ker(G) = \ker(\overline{G}) = U$*. Then the codes* $\mathcal{C}_{U,G}$ *and* $\mathcal{C}_{U,\overline{G}}$ *are equivalent.*

Proof. Let $B_U \cup \{w_1, \ldots, w_{rm-k}\}$ be an $\mathbb{F}_q$-basis of V such that B_U is an $\mathbb{F}_q$-basis of U. Clearly, since G and $\overline{G}$ are $\mathbb{F}_q$-linear maps, $G(w_1), \ldots, G(w_{rm-k})$ are $\mathbb{F}_q$-linearly independent, as well as $\overline{G}(w_1), \ldots, \overline{G}(w_{rm-k})$. Then there exists an invertible $\mathbb{F}_q$-linear map $L \colon W \to W$ such that $L(G(w_i)) = \overline{G}(w_i)$ for every $i = 1, \ldots, rm-k$, i.e. $L \circ G = \overline{G}$. Therefore, by choosing $R = \mathrm{Id}_{\mathbb{F}_{q^m}}$ and $\sigma = \mathrm{Id}_{\mathrm{Aut}(\mathbb{F}_q)}$, we have

$$L \circ \mathcal{C}_{U,G}^{\sigma} \circ R = \{ L \circ (G \circ \tau_{\mathbf{v}}) : \mathbf{v} \in V \} = \mathcal{C}_{U,\overline{G}},$$

which completes the proof. ∎

Not only can we construct $\mathbb{F}_q$-linear MRD codes using Theorem 3.4.3, but, in fact, any $\mathbb{F}_q$-linear MRD code as in the claim of that theorem is equivalent to a code $\mathcal{C}_{U,G}$ for a certain subspace U. Because the proof of this statement relies on Singer cycles of $\mathrm{GL}_m(q)$, we first provide the definition thereof and state a lemma (without detailed proof), before proving the converse statement.

Definition 3.4.5 (Singer cycle). The cyclic subgroups of $\mathrm{GL}_m(q)$ of order $q^m - 1$ are called Singer cycles.

Lemma 3.4.6. *Let* S_1 *and* S_2 *be two Singer cycles of* $\mathrm{GL}_m(q)$*. Then* S_1 *and* S_2 *are conjugate in* $\mathrm{GL}_m(q)$*. That is, there exists an invertible* $\mathbb{F}_q$*-linear map* $\psi \colon \mathbb{F}_{q^m} \to \mathbb{F}_{q^m}$ *such that* $S_2 = \psi \circ S_1 \circ \psi^{-1}$*.*

Proof. See, for example, [41, Section 4.2]. ∎

Theorem 3.4.7 ([104, Theorem 3.5]). *Let* $\mathcal{C}$ *be an* $\mathbb{F}_q$*-linear MRD code with parameters* $(m, t, q; m-\iota)$ *such that* $t \geq m$ *and* $|L(\mathcal{C})| = q^m$*, contained in* $\mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_{q^m}, W)$ *with* $W = V(t, q)$*. Let* $r = \dim_{L(\mathcal{C})}(\mathcal{C})$*. Then the following holds:*

(i) $\iota + 1 \mid rm$ *and* $t = \frac{rm}{\iota+1}$*;*

(ii) $\mathcal{C}$ *is equivalent to an* $\mathbb{F}_q$*-linear MRD code* $\mathcal{C}'$ *such that* $L(\mathcal{C}') = \mathcal{F}_m$*;*

(iii) The set

$$U = \left\{ f \in \mathcal{C}' : f(1) = 0 \right\} \subseteq \mathcal{C}'$$

is an $\frac{\iota rm}{\iota+1}$*-dimensional* $\mathcal{F}_{m,q}$*-subspace of* $\mathcal{C}'$*, and satisfies*

$$\max\left\{\dim_{\mathcal{F}_{m,q}}(U \cap \langle f \rangle_{\mathcal{F}_m}) : f \in \mathcal{C}' \setminus \{0\}\right\} = \iota,$$

where $\langle f \rangle_{\mathcal{F}_m} = \{ f \circ \omega_\alpha : \alpha \in \mathbb{F}_{q^m} \}$*;*

(iv) $\mathcal{C}'$ *is equivalent to* $\mathcal{C}_{U,G}$*, where* $G \colon \mathcal{C}' \to W, f \mapsto f(1)$*.*

Proof. Because $|L(\mathcal{C})| = q^m$ and $r = \dim_{L(\mathcal{C})}(\mathcal{C})$, we have $|\mathcal{C}| = q^{rm}$. In combination with the assumption $t \geq m$, the Singleton-like bound in Equation (2.2) reads $rm \leq t(m - (m - \iota) + 1)$. As $\mathcal{C}$ is MRD, this implies that $\iota + 1$ divides rm, and $t = \frac{rm}{\iota+1}$, proving (i).

For (ii), since $L(\mathcal{C})$ and $\mathcal{F}_m$ are isomorphic to fields of order q^m (for $L(\mathcal{C})$, see Theorem 3.1.12), their multiplicative groups are cyclic of order $q^m - 1$; hence, they define Singer cycles in $\mathrm{GL}_m(q)$. By Lemma 3.4.6, there exists an invertible $\mathbb{F}_q$-linear map $\psi \colon \mathbb{F}_{q^m} \to \mathbb{F}_{q^m}$ such that $L(\mathcal{C}) = \psi \circ \mathcal{F}_m \circ \psi^{-1}$.

We now define $\mathcal{C}' = \mathcal{C} \circ \psi$. Since ψ is an invertible $\mathbb{F}_q$-linear map, $\mathcal{C}'$ is an $\mathbb{F}_q$-linear MRD code equivalent to $\mathcal{C}$. Moreover, for any $\omega_\alpha \in \mathcal{F}_m$ and $h = f \circ \psi \in \mathcal{C}'$, $f \in \mathcal{C}$, we observe that $h \circ \omega_\alpha = f \circ (\psi \circ \omega_\alpha \circ \psi^{-1}) \circ \psi$. Because $\psi \circ \omega_\alpha \circ \psi^{-1} \in L(\mathcal{C})$, the composition $f' = f \circ (\psi \circ \omega_\alpha \circ \psi^{-1})$ remains in $\mathcal{C}$, implying $h \circ \omega_\alpha = f' \circ \psi \in \mathcal{C}'$ and confirming $L(\mathcal{C}') = \mathcal{F}_m$.

Next, we prove (iii). The set U, being the kernel of the $\mathbb{F}_q$-linear evaluation map $f \mapsto f(1)$ defined on the $\mathbb{F}_q$-linear code $\mathcal{C}'$, forms an $\mathcal{F}_{m,q}$-subspace (i.e. an $\mathbb{F}_q$-subspace) of $\mathcal{C}'$. Now, define, for every $i \in \{1, \ldots, \iota\}$, the set

$$U_i = \left\{ f \in U : \dim_{\mathbb{F}_q}(\ker(f)) = i \right\}.$$

Let $g \in \mathcal{C}'$ be such that $\dim_{\mathbb{F}_q}(\ker(g)) = i$. We know that $\mathcal{C}'$ is a left vector space over $\mathcal{F}_m$ because its left idealiser is $L(\mathcal{C}') = \mathcal{F}_m$, ensuring closure under pre-composition by scalars. For any non-zero $\alpha \in \ker(g)$, we have $g(\alpha) = 0$, meaning $(g \circ \omega_\alpha)(1) = 0$. Thus, by setting $f = g \circ \omega_\alpha$, we ensure $f \in \mathcal{C}'$ and $1 \in \ker(f)$. In particular, we have $f \in U_i$. By rewriting $g = f \circ \omega_{\alpha^{-1}}$, this implies that the set

$$\{ f \circ \omega_\beta : f \in U_i, \beta \in \mathbb{F}_{q^m} \setminus \{0\} \}$$

generates every element $g \in \mathcal{C}'$ of rank $m - i$. Note that a pair $(f, \beta) = (g \circ \omega_{\beta^{-1}}, \beta)$ maps to the same function $g = f \circ \omega_\beta$ if and only if $\beta^{-1} \in \ker(g)$. A double-counting argument then gives

$$A_{m-i} = \frac{|U_i|(q^m - 1)}{q^i - 1}.$$

Furthermore, for any non-zero $f \in \mathcal{C}'$, the intersection $U \cap \langle f \rangle_{\mathcal{F}_m}$ consists of the elements $f \circ \omega_\gamma$ satisfying $(f \circ \omega_\gamma)(1) = f(\gamma) = 0$. Thus, $\dim_{\mathcal{F}_{m,q}}(U \cap \langle f \rangle_{\mathcal{F}_m}) = \dim_{\mathbb{F}_q}(\ker(f))$. By Lemma 3.4.2, we know $A_{m-\iota} \neq 0$, meaning the code contains at least one codeword achieving the minimum possible non-zero rank $m - \iota$. From this, $\max\left\{\dim_{\mathcal{F}_{m,q}}(U \cap \langle f \rangle_{\mathcal{F}_m}) : f \in \mathcal{C}' \setminus \{0\}\right\} = \iota$ follows.

The cardinality of U is determined by partitioning it into the zero vector and the sets U_i. Using the expression for $|U_i|$ derived above, we write

$$|U| = 1 + \sum_{i=1}^{\iota} |U_i| = 1 + \frac{1}{q^m - 1} \sum_{i=1}^{\iota} A_{m-i}(q^i - 1).$$

Multiplying by $q^m - 1$ gives the relation

$$(|U| - 1)(q^m - 1) = \sum_{i=1}^{\iota} A_{m-i}(q^i - 1).$$

By Theorem 3.1.8 applied to $\mathcal{C}'$ with $\nu = 1$, we find that the right-hand side equals $(q^m - 1)(q^{\frac{\iota rm}{\iota+1}} - 1)$. Substituting this back simplifies the equation to

$$|U| = q^{\frac{\iota rm}{\iota+1}}.$$

Since U is an $\mathbb{F}_q$-vector space, its dimension is the exponent of q; hence, $\dim_{\mathcal{F}_{m,q}}(U) = \frac{\iota rm}{\iota+1}$.

Finally, we show that $\mathcal{C}'$ coincides with the construction $\mathcal{C}_{U,G}$ by setting $G\colon \mathcal{C}' \to W$ to be the evaluation map $f \mapsto f(1)$ (note that $\ker(G) = U$). By definition, the codewords of $\mathcal{C}_{U,G}$ are the maps $\Gamma_f = G \circ \tau_f$, where $\tau_f(\alpha) = f \circ \omega_\alpha$ represents scalar multiplication in $\mathcal{C}'$. For any $\alpha \in \mathbb{F}_{q^m}$, we have

$$\Gamma_f(\alpha) = G(f \circ \omega_\alpha) = (f \circ \omega_\alpha)(1) = f(\alpha).$$

Since Γ_f acts identically to f, we conclude $\mathcal{C}_{U,G} = \{\, \Gamma_f : f \in \mathcal{C}' \,\} = \mathcal{C}'$, proving (iv). ∎

So, in conclusion, Theorem 3.4.3 and Theorem 3.4.7 provide a correspondence between, on the one hand, $\mathbb{F}_q$-subspaces $U = V(\frac{\iota rm}{\iota+1}, q)$ of $V = V(r, q^m)$ such that $\iota = \max\{\dim_{\mathbb{F}_q}(U \cap \langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}}) : \mathbf{v} \in V \setminus \{\mathbf{0}\}\}$ and, on the other hand, $\mathbb{F}_q$-linear MRD codes $\mathcal{C}$ with parameters $\left(m, \frac{rm}{\iota+1}, q; m - \iota\right)$ and left idealiser isomorphic to $\mathbb{F}_{q^m}$.

3.4.2 MRD codes obtained via maximum h-scattered subspaces

At this point, we know how to build a rank-metric code $\mathcal{C}_{U,G}$ from a subspace U. However, it remains unclear which U to choose such that the resulting code is MRD. Fortunately, a link with h-scattered subspaces not only clears the fog on this matter, but also allows us to prove inequivalence with existing codes for some constructions and helps us to determine its rank distribution.

Definition 3.4.8 (h-scattered subspace). An $\mathbb{F}_q$-subspace U of an r-dimensional $\mathbb{F}_{q^m}$-vector space V is said to be *h-scattered*, with $h \leq r - 1$ a positive integer, if U spans V over $\mathbb{F}_{q^m}$ and, for any h-dimensional $\mathbb{F}_{q^m}$-subspace H of V, U meets H in an $\mathbb{F}_q$-subspace of dimension at most h.

An h-scattered $\mathbb{F}_q$-subspace of highest dimension in $V(r, q^m)$ is called *maximum h-scattered*; its dimension is upper bounded by the following proposition.

Proposition 3.4.9 ([19, Theorem 2.3]). *If U is an h-scattered $\mathbb{F}_q$-subspace of dimension k in $V = V(r, q^m)$, then one of the following holds:*

- *$k = r$ and U defines a subgeometry* $\mathrm{PG}(r-1, q)$ *of* $\mathrm{PG}(V, \mathbb{F}_{q^m})$*;*
- *$k \leq \frac{rm}{h+1}$.*

The seminal paper [94] by Sheekey initially established the link between MRD codes and scattered $\mathbb{F}_q$-subspaces in the context of maximum 1-scattered subspaces within $V(2, q^m)$. This correspondence was subsequently extended in [20] to maximum 1-scattered $\mathbb{F}_q$-subspaces in general dimension $V(r, q^m)$ and further analysed by Polverino and Zullo [83]. Simultaneously, connections regarding the other extremal case, namely maximum $(r-1)$-scattered subspaces in $V(r, q^m)$, were identified in [61, 96]. Here, we present the work of Zini and Zullo [104] that bridged the gap between these known instances ($h = 1$ and $h = r - 1$) by proposing a unified framework valid for any $h \in \{1, \ldots, r-1\}$.

The main observation is that the required U's for the code $\mathcal{C}_{U,G}$ to be MRD can be constructed as the ordinary duals of $\frac{rm}{h+1}$-dimensional, i.e. maximum, h-scattered $\mathbb{F}_q$-subspaces of $V = V(r, q^m)$. A key ingredient for this duality is the trace of a finite field.

Definition 3.4.10 (Trace). For $\alpha \in \mathbb{F}_{q^m}$, the *trace* $\mathrm{Tr}_{q^m/q}(\alpha)$ of α over the subfield $\mathbb{F}_q$ is defined by

$$\mathrm{Tr}_{q^m/q}(\alpha) = \sum_{i=0}^{m-1} \alpha^{q^i} = \alpha + \alpha^q + \cdots + \alpha^{q^{m-1}}.$$

When $\mathbb{F}_q$ is the prime subfield, it is called the *absolute trace* and simply denoted by $\mathrm{Tr}_{\mathbb{F}_q}(\alpha)$.

In fact, our treatment of the problem requires two types of duality for $\mathbb{F}_q$-subspaces: ordinary and Delsarte duality.

Ordinary duality. The ordinary duality makes use of a non-degenerate reflexive sesquilinear form $\sigma\colon V \times V \to \mathbb{F}_{q^m}$ over $V = V(r, q^m)$. We define $\sigma'\colon V \times V \to \mathbb{F}_q, (u, v) \mapsto \mathrm{Tr}_{q^m/q}(\sigma(u, v))$ and observe that σ' is a non-degenerate reflexive sesquilinear form over $V = V(rm, q)$. Let $\perp$ and $\perp'$ be the orthogonal complement maps defined by σ and σ', respectively, on the lattices of $\mathbb{F}_{q^m}$-subspaces and $\mathbb{F}_q$-subspaces of V. When U is an $\mathbb{F}_q$-subspace of V, we denote by $U^{\perp_O}$ one of the $\mathbb{F}_q$-subspaces $U^{\perp'}$ (corresponding to any non-degenerate reflexive sesquilinear form σ').

The following properties will be useful later.

Proposition 3.4.11 ([82, Section 2]). *Let $V = V(r, q^m)$ and $\perp, \perp'$ be as defined above. Then we have:*

(i) $\dim_{\mathbb{F}_{q^m}}(W) + \dim_{\mathbb{F}_{q^m}}(W^{\perp}) = r$, *for every $\mathbb{F}_{q^m}$-subspace W of V;*

(ii) $\dim_{\mathbb{F}_q}(U) + \dim_{\mathbb{F}_q}(U^{\perp'}) = mr$, *for every $\mathbb{F}_q$-subspace U of V;*

(iii) $W^{\perp} = W^{\perp'}$ *for every $\mathbb{F}_{q^m}$-subspace W of V;*

(iv) Let W and U be an $\mathbb{F}_{q^m}$-subspace and an $\mathbb{F}_q$-subspace of V of dimensions s and t, respectively, then

$$\dim_{\mathbb{F}_q}(U^{\perp'} \cap W^{\perp'}) - \dim_{\mathbb{F}_q}(U \cap W) = rm - \dim_{\mathbb{F}_q}(U) - sm.$$

Delsarte duality. To define the Delsarte duality, we follow [19, Section 3]. Firstly, we require an embedding of our ambient space into a larger space $\mathbb{V}$. Let U be a k-dimensional $\mathbb{F}_q$-subspace of $V = V(r, q^m)$ such that it spans V over $\mathbb{F}_{q^m}$, i.e. $\langle U\rangle_{\mathbb{F}_{q^m}} = V$, and assume $k > r$. We consider a space $\mathbb{V} = V(k, q^m)$ which allows for the decomposition $\mathbb{V} = V \oplus \Gamma$ (component-wise operations), where Γ is an $\mathbb{F}_{q^m}$-subspace of dimension $k - r$.

The subspace U can be represented as the restriction of a subspace W from this larger space. By [64, Theorem 2], there exists a k-dimensional $\mathbb{F}_q$-subspace W of $\mathbb{V}$ such that $\langle W\rangle_{\mathbb{F}_{q^m}} = \mathbb{V}$, $W \cap \Gamma = \{0\}$, and

$$U = \langle W, \Gamma\rangle_{\mathbb{F}_q} \cap V.$$

Let $\beta'\colon W \times W \to \mathbb{F}_q$ be a non-degenerate reflexive sesquilinear form. This can be extended to a form β on $\mathbb{V}$. Let $\perp_{\mathbb{V}}$ denote the orthogonal complement map defined by β on the lattice of $\mathbb{F}_{q^m}$-subspaces of $\mathbb{V}$.

Definition 3.4.12 (Delsarte dual). Let U be a k-dimensional $\mathbb{F}_q$-subspace of $V = V(r, q^m)$ such that $k > r$ and $\dim_{\mathbb{F}_q}(M \cap U) < k - 1$ for every $(r-1)$-dimensional $\mathbb{F}_{q^m}$-subspace M of V. Then the k-dimensional $\mathbb{F}_q$-subspace $W + \Gamma^{\perp_{\mathbb{V}}}$ of the quotient space $\mathbb{V}/\Gamma^{\perp_{\mathbb{V}}}$ will be denoted by $U^{\perp_D}$ and will be called the Delsarte dual of U (w.r.t. $\perp_{\mathbb{V}}$).

It is worth noting that while U resides in $V(r, q^m)$, its Delsarte dual $U^{\perp_D}$ resides in a space isomorphic to $V(k - r, q^m)$. The Delsarte dual is, in particular, useful in the context of scattered subspaces, as the following theorem shows.

Theorem 3.4.13 ([19, 104]). *Let $U, V, \Gamma, \mathbb{V}, \perp_{\mathbb{V}}$, and $\perp_D$ be defined as above. Under the assumption $m \geq h + 3$, U is an $\frac{rm}{h+1}$-dimensional h-scattered $\mathbb{F}_q$-subspace of V if and only if $U^{\perp_D}$ is an $\frac{rm}{h+1}$-dimensional $(m - h - 2)$-scattered $\mathbb{F}_q$-subspace of $\mathbb{V}/\Gamma^{\perp_{\mathbb{V}}}$.*

In the remainder of this section, we follow [104] and prove that, whenever $m \geq h + 3$, the required subspaces U to construct MRD codes are exactly the ordinary duals of $\frac{rm}{h+1}$-dimensional h-scattered $\mathbb{F}_q$-subspaces of V. To this aim, we characterise maximum h-scattered subspaces.

Theorem 3.4.14 ([104, Theorem 5.1]). *Let r, m, h, k be positive integers such that $m \geq h+3$ and $k > r$. Let U be a k-dimensional $\mathbb{F}_q$-subspace of $V = V(r, q^m)$ such that*

$$\dim_{\mathbb{F}_q}(H \cap U) \leq k - m + h \tag{3.6}$$

for every $(r-1)$-dimensional $\mathbb{F}_{q^m}$-subspace H of V. Let $\Gamma, \mathbb{V}, \perp_{\mathbb{V}}, \perp_D$ be as defined above. Then $U^{\perp_D}$ is an $(m-h-2)$-scattered subspace of $\mathbb{V}/\Gamma^{\perp_{\mathbb{V}}}$.

Proof. As noted in the part about Delsarte duality above, there exist $\mathbb{V} = V(k, q^m)$, an $\mathbb{F}_{q^m}$-subspace $\Gamma = V(k-r, q^m)$ of $\mathbb{V}$, and an $\mathbb{F}_q$-subspace $W = V(k, q)$ of $\mathbb{V}$ such that $\mathbb{V} = V \oplus \Gamma$, $\langle W \rangle_{\mathbb{F}_{q^m}} = \mathbb{V}$, $W \cap \Gamma = \{0\}$, and $U = \langle W, \Gamma \rangle_{\mathbb{F}_q} \cap V$.

Let $\perp'_{\mathbb{V}}$ and $\perp_{\mathbb{V}}$ be the orthogonal complement maps which act, respectively, on the $\mathbb{F}_q$-subspaces of W and on the $\mathbb{F}_{q^m}$-subspaces of $\mathbb{V}$. They are defined by non-degenerate reflexive sesquilinear forms $\beta' \colon W \times W \to \mathbb{F}_q$ and $\beta \colon \mathbb{V} \times \mathbb{V} \to \mathbb{F}_{q^m}$ respectively, such that $\beta\big|_{W \times W} = \beta'$.

Since $m \geq h+3$, $k > r$, and Equation (3.6) holds, the Delsarte duality can be applied to U. Note that, strictly speaking, $m \geq h+2$ suffices, but the result then becomes trivial (0-scattered subspace). So, $U^{\perp_D} = W + \Gamma^{\perp_{\mathbb{V}}}$ is a k-dimensional $\mathbb{F}_q$-subspace of $\mathbb{V}/\Gamma^{\perp_{\mathbb{V}}}$.

Suppose, for the sake of contradiction, that there exists an $(m-h-2)$-dimensional $\mathbb{F}_{q^m}$-subspace M of $\mathbb{V}/\Gamma^{\perp_{\mathbb{V}}}$ such that $\dim_{\mathbb{F}_q}(M \cap U^{\perp_D}) \geq m-h-1$. Write $M = N/\Gamma^{\perp_{\mathbb{V}}}$, where N is an $(m-h-2+r)$-dimensional subspace of $\mathbb{V}$ satisfying $\Gamma^{\perp_{\mathbb{V}}} \subseteq N$, so that

$$\dim_{\mathbb{F}_q}(N \cap W) = \dim_{\mathbb{F}_q}(M \cap U^{\perp_D}) \geq m - h - 1.$$

This equality holds because W intersects the kernel $\Gamma^{\perp_{\mathbb{V}}}$ trivially (a consequence of the rank-nullity theorem, since $\dim_{\mathbb{F}_q}(W) = \dim_{\mathbb{F}_q}(U^{\perp_D})$). This implies that the canonical quotient map $\mathbb{V} \to \mathbb{V}/\Gamma^{\perp_{\mathbb{V}}}$ restricts to a linear isomorphism from W to $U^{\perp_D}$, mapping $N \cap W$ bijectively onto $M \cap U^{\perp_D}$.

From now on, denote the $\mathbb{F}_{q^m}$-subspace $\langle S \rangle_{\mathbb{F}_{q^m}}$ of $\mathbb{V}$, for an $\mathbb{F}_q$-subspace S of W, as S^*. Take for S an $(m-h-1)$-dimensional $\mathbb{F}_q$-subspace of $N \cap W$. As $S \subseteq W$, we have $\dim_{\mathbb{F}_{q^m}}(S^*) = \dim_{\mathbb{F}_q}(S)$. Indeed, since $S \subseteq W$ and W is a canonical subgeometry of $\mathbb{V}$ (meaning any $\mathbb{F}_q$-basis of W is an $\mathbb{F}_{q^m}$-basis of $\mathbb{V}$), the $\mathbb{F}_q$-basis of S remains linearly independent over $\mathbb{F}_{q^m}$, forming a basis for S^*; see [62, Lemma 1]. Since N contains both S and $\Gamma^{\perp_{\mathbb{V}}}$, we have $N^{\perp_{\mathbb{V}}} \subseteq (S^*)^{\perp_{\mathbb{V}}} \cap \Gamma$, whence

$$\dim_{\mathbb{F}_{q^m}}((S^*)^{\perp_{\mathbb{V}}} \cap \Gamma) \geq \dim_{\mathbb{F}_{q^m}}(N^{\perp_{\mathbb{V}}}) = k - (m - h - 2 + r).$$

This implies that $\langle (S^*)^{\perp_{\mathbb{V}}}, \Gamma \rangle_{\mathbb{F}_{q^m}}$ is contained in an $\mathbb{F}_{q^m}$-subspace T of $\mathbb{V}$ of dimension $k-1$, due to Grassmann's identity:

$$\begin{aligned}
\dim_{\mathbb{F}_{q^m}}(T) &= \dim_{\mathbb{F}_{q^m}}((S^*)^{\perp_{\mathbb{V}}}) + \dim_{\mathbb{F}_{q^m}}(\Gamma) - \dim_{\mathbb{F}_{q^m}}((S^*)^{\perp_{\mathbb{V}}} \cap \Gamma) \\
&\leq (k - (m-h-1)) + (k-r) - (k - (m-h-2+r)) \\
&\leq k - 1.
\end{aligned}$$

Let $\hat{T} = T \cap V$. As T contains Γ and is a subspace of $\mathbb{V} = V \oplus \Gamma$, it holds that $\dim_{\mathbb{F}_{q^m}}(\hat{T}) = \dim_{\mathbb{F}_{q^m}}(T) - \dim_{\mathbb{F}_{q^m}}(\Gamma) = k - 1 - (k-r) = r-1$. Using $U = \langle W, \Gamma \rangle_{\mathbb{F}_q} \cap V$ and $T \cap \langle W, \Gamma \rangle_{\mathbb{F}_q} = \langle \Gamma, T \cap W \rangle_{\mathbb{F}_q}$ (because $\Gamma \subseteq T$), we obtain:

$$\begin{aligned}
\dim_{\mathbb{F}_q}(\hat{T} \cap U) &= \dim_{\mathbb{F}_q}(\langle \Gamma, T \cap W \rangle_{\mathbb{F}_q} \cap V) \\
&= \dim_{\mathbb{F}_q}(\langle \Gamma, T \cap W \rangle_{\mathbb{F}_q}) + \dim_{\mathbb{F}_q}(V) - \dim_{\mathbb{F}_q}(\langle \langle \Gamma, T \cap W \rangle_{\mathbb{F}_q}, V \rangle_{\mathbb{F}_q}) \\
&= m(k-r) + \dim_{\mathbb{F}_q}(T \cap W) + mr - \dim_{\mathbb{F}_q}(\mathbb{V}) \\
&= mk - mr + \dim_{\mathbb{F}_q}(T \cap W) + mr - mk \\
&= \dim_{\mathbb{F}_q}(T \cap W).
\end{aligned}$$

The third equality follows because $\Gamma \cap W = \{0\}$. As $S^{\perp'_{\mathbb{V}}} = W \cap (S^*)^{\perp_{\mathbb{V}}} \subseteq W \cap T$ and $\dim_{\mathbb{F}_q}(S^{\perp'_{\mathbb{V}}}) = k - (m - h - 1)$, we obtain

$$\dim_{\mathbb{F}_q}(\hat{T} \cap U) \geq k - m + h + 1,$$

which contradicts our assumption (3.6). Therefore, we conclude that $U^{\perp_D}$ is an $(m - h - 2)$-scattered $\mathbb{F}_q$-subspace of $\mathbb{V}/\Gamma^{\perp_{\mathbb{V}}}$. ∎

The previous theorem establishes that a specific bound on the intersection dimension with hyperplanes is sufficient to ensure the Delsarte dual is scattered. To move towards a complete characterisation, we must determine whether maximum h-scattered subspaces naturally satisfy this bound. The following lemma clarifies the intersection properties of h-scattered subspaces, specifically when they attain the maximum dimension $\frac{rm}{h+1}$.

Lemma 3.4.15 ([19, Theorem 2.7]). *If U is an $\frac{rm}{h+1}$-dimensional h-scattered $\mathbb{F}_q$-subspace of a vector space $V = V(r, q^m)$, then for any $(r-1)$-dimensional $\mathbb{F}_{q^m}$-subspace H of V we have*

$$\frac{rm}{h+1} - m \leq \dim_{\mathbb{F}_q}(U \cap H) \leq \frac{rm}{h+1} - m + h.$$

Proof. [19, Section 5] is devoted to the proof of this lemma. ∎

Note that the upper bound derived in Lemma 3.4.15 coincides exactly with the condition in Theorem 3.4.14 when we set $k = \frac{rm}{h+1}$. This observation provides the following characterisation.

Corollary 3.4.16 ([104, Corollary 5.2]). *Let r, m, h be positive integers such that $h+1$ divides rm and $m \geq h+3$. Let U be an $\frac{rm}{h+1}$-dimensional $\mathbb{F}_q$-subspace of $V = V(r, q^m)$. Then U is an $\frac{rm}{h+1}$-dimensional h-scattered $\mathbb{F}_q$-subspace of V if and only if*

$$\dim_{\mathbb{F}_q}(H \cap U) \leq \frac{rm}{h+1} - m + h \tag{3.7}$$

for every $(r-1)$-dimensional $\mathbb{F}_{q^m}$-subspace H of V.

Proof. Assume that (3.7) holds. By Theorem 3.4.14, $U^{\perp_D}$ is an $\frac{rm}{h+1}$-dimensional $(m-h-2)$-scattered $\mathbb{F}_q$-subspace in $\mathbb{V}/\Gamma^{\perp_{\mathbb{V}}}$. By Theorem 3.4.13, U is an $\frac{rm}{h+1}$-dimensional h-scattered $\mathbb{F}_q$-subspace of V. The converse follows from Lemma 3.4.15. ∎

Corollary 3.4.16 cannot be extended to all h-scattered subspaces of arbitrary dimension; it only holds for maximum h-scattered subspaces, that is, $\frac{rm}{h+1}$-dimensional subspaces. This is shown in the next counterexample, which is inspired by [104, Remark 5.3].

Example 3.4.17. Let $r = 3$, $m = 4$, and $h = 1$. The maximum dimension for a 1-scattered subspace in $V = V(3, q^4)$ is $k_{\max} = (3 \cdot 4)/2 = 6$. We construct an $\mathbb{F}_q$-subspace U of dimension $k = 5$ that is 1-scattered but does not satisfy the condition (3.6).

Let H be a 2-dimensional $\mathbb{F}_{q^m}$-subspace of V (a hyperplane). Let U' be a maximum 1-scattered $\mathbb{F}_q$-subspace of H. Since $H \cong V(2, q^4)$, the dimension of U' is $k' = (2 \cdot 4)/2 = 4$. By definition, U' spans H over $\mathbb{F}_{q^m}$, so $\langle U' \rangle_{\mathbb{F}_{q^m}} = H$.

Now, choose a vector $\mathbf{v} \in V \setminus H$ and define the $\mathbb{F}_q$-subspace

$$U = U' \oplus \langle \mathbf{v} \rangle_{\mathbb{F}_q}.$$

The dimension of U is $k = k' + 1 = 5$. Since $\mathbf{v} \notin H$, the space U spans V over $\mathbb{F}_{q^m}$. Furthermore, U is 1-scattered: any 1-dimensional $\mathbb{F}_{q^m}$-subspace contained in H intersects U in at most 1 dimension (since U' is scattered), and any 1-dimensional $\mathbb{F}_{q^m}$-subspace not contained in H intersects U in at most 1 dimension over $\mathbb{F}_q$ (due to the direct sum with the 1-dimensional space $\langle \mathbf{v} \rangle_{\mathbb{F}_q}$).

However, let us test the intersection condition for the specific hyperplane H. The condition requires that, for every hyperplane,

$$\dim_{\mathbb{F}_q}(H \cap U) \leq k - m + h = 5 - 4 + 1 = 2.$$

Because $\mathbf{v} \notin H$, the intersection of U and H is exactly U'. Therefore,

$$\dim_{\mathbb{F}_q}(H \cap U) = \dim_{\mathbb{F}_q}(U') = 4.$$

Since $4 \not\leq 2$, U violates the intersection bound. This shows that the characterisation in Corollary 3.4.16 cannot be extended to non-maximum 1-scattered subspaces, i.e. when $k < k_{\max} = \frac{rm}{h+1}$. □

We can also find a characterisation based on ordinary duality, using Corollary 3.4.16.

Corollary 3.4.18 ([104, Corollary 5.4]). *Let r, m, h be positive integers such that $h+1$ divides rm and $m \geq h+3$. Let U be an $\frac{rm}{h+1}$-dimensional $\mathbb{F}_q$-subspace of $V = V(r, q^m)$. Then U is an $\frac{rm}{h+1}$-dimensional h-scattered $\mathbb{F}_q$-subspace of V if and only if $U^{\perp_O}$ satisfies*

$$\dim_{\mathbb{F}_q}(\langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}} \cap U^{\perp_O}) \leq h \tag{3.8}$$

for every $\mathbf{v} \in V \setminus \{\mathbf{0}\}$.

Proof. First of all, assume that U is an $\frac{rm}{h+1}$-dimensional h-scattered $\mathbb{F}_q$-subspace of V. Define $W^{\perp'} = \langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}}$. Then $\dim_{\mathbb{F}_{q^m}}(W^{\perp'}) = 1$ and, by Proposition 3.4.11(i) and (iii), we get $\dim_{\mathbb{F}_{q^m}}(W) = r - 1$. Therefore, using Lemma 3.4.15 and Proposition 3.4.11(iv), we find

$$\begin{aligned}
\dim_{\mathbb{F}_q}(\langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}} \cap U^{\perp_O}) &= rm - \dim_{\mathbb{F}_q}(U) - (r-1)m + \dim_{\mathbb{F}_q}(U \cap W) \\
&\leq -\frac{rm}{h+1} + m + \frac{rm}{h+1} - m + h \\
&\leq h,
\end{aligned}$$

because $U^{\perp_O}$ is a subspace $U^{\perp'}$.

Conversely, $\dim_{\mathbb{F}_q}(U) = \frac{rm}{h+1}$ implies $\dim_{\mathbb{F}_q}(U^{\perp_O}) = rm - \frac{rm}{h+1} = \frac{hrm}{h+1}$. Together with assumption (3.8) and Proposition 3.4.11(iv), this yields, for every $(r-1)$-dimensional $\mathbb{F}_{q^m}$-subspace H of V,

$$\begin{aligned}
\dim_{\mathbb{F}_q}(H \cap U) &= (r-1)m + \dim_{\mathbb{F}_q}(U) - rm + \dim_{\mathbb{F}_q}(H^{\perp'} \cap U^{\perp'}) \\
&= -m + \frac{rm}{h+1} + \dim_{\mathbb{F}_q}(\langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}} \cap U^{\perp_O}) \\
&\leq \frac{rm}{h+1} - m + h.
\end{aligned}$$

The claim now follows from Corollary 3.4.16. ■

Observe that the intersection condition on the ordinary dual $U^{\perp_O}$ in Corollary 3.4.18 and its dimension $\frac{hrm}{h+1}$ are exactly the conditions required by Theorem 3.4.3 and Theorem 3.4.7 to obtain an MRD code (by setting $\iota = h$). The ordinary dual of a maximum h-scattered subspace U can therefore be used to construct the code $\mathcal{C}_{U^{\perp_O},G}$.

In conclusion, under the assumption $m \geq h+3$, there is a correspondence between:

- $\frac{rm}{h+1}$-dimensional h-scattered $\mathbb{F}_q$-subspaces of $V(r, q^m)$; and
- $\mathbb{F}_q$-linear MRD codes with parameters $\left(m, \frac{rm}{h+1}, q; m-h\right)$ and left idealiser isomorphic to $\mathbb{F}_{q^m}$.

3.4.3 Inequivalence with Gabidulin codes

At this point, the question remains whether this new construction yields codes that are inequivalent to known families, such as the generalised additive twisted Gabidulin codes. We can immediately rule out equivalence with purely additive codes: because linearity is preserved under code equivalence and the constructed MRD codes are $\mathbb{F}_q$-linear (see Theorem 3.4.3), they cannot be equivalent to any non-linear code.

We will now prove that, when $(h+1) \nmid r$, the constructed MRD code using Equation (3.5) is inequivalent to a generalised (twisted) Gabidulin code as defined in Definition 3.3.8, or any punctured version of it.

Definition 3.4.19 (Punctured code). Let $\mathcal{C} \subseteq \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^n)$ be a rank-metric code. Let $k < n$ be a positive integer and let $\rho \in \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^n, \mathbb{F}_q^k)$ be a surjective linear map. The *punctured code* $\mathcal{C}_\rho$ is defined as the set of compositions

$$\mathcal{C}_\rho = \{\rho \circ f : f \in \mathcal{C}\} \subseteq \mathrm{Hom}_{\mathbb{F}_q}(\mathbb{F}_q^m, \mathbb{F}_q^k).$$

Remark 3.4.20. In terms of matrix representations, we fix bases such that ρ is the canonical projection and its matrix representation is the $n \times k$ block matrix $A = \left(I_k \mid \mathbf{0}_{n-k}\right)^\top$, where I_k denotes the $k \times k$ identity matrix and $\mathbf{0}_{n-k}$ the $k \times (n-k)$ null matrix. The matrices of $\mathcal{C}_\rho$ are obtained by multiplying the matrices in $\mathcal{C}$ on the right by A:

$$\mathcal{C}_\rho = \{\, CA : C \in \mathcal{C} \,\}.$$

This operation deletes the last $n-k$ columns of the matrices in $\mathcal{C}$.

It is important to consider the operation of puncturing, because – as hypothesised by Sheekey [94] and proven by Csajbók and Siciliano [17] – puncturing generalised twisted Gabidulin codes can yield MRD codes that are inequivalent to the original generalised Gabidulin family. However, it was shown by Zini and Zullo [104] that the MRD codes from Section 3.4.1 do not merely replicate these punctured structures, but are genuinely new codes. For this purpose, we need the following lemma.

Lemma 3.4.21 ([104, Lemma 6.1]). *Let $\rho\colon \mathbb{F}_{q^M} \to \mathbb{F}_{q^N}$ be an $\mathbb{F}_q$-linear map of rank $N \leq M$, and consider the punctured code $\mathcal{D}$, where either $\mathcal{D} = \rho \circ \mathcal{G}_{k,\sigma}$ or $\mathcal{D} = \rho \circ \mathcal{H}_{k,\sigma}(\eta, c)$. If $N > k+1$, and $(N,k) \neq (4,2)$, then $|R(\mathcal{C})| = q^\ell$ where $\ell \mid M$.*

Proof. This is a consequence of [101, Theorem 3.8]. ∎

Next, we need the concept of the adjoint code, most easily described in the matrix representation. This is because the code given by Equation (3.5) has parameters $(m,n,q;d)$ with $m \leq n$, while the punctured $(m,n,q;d)$-codes $\mathcal{C}_\rho$ arising from Lemma 3.4.21 satisfy $m \geq n$.

Definition 3.4.22 (Adjoint code). For a rank-metric code $\mathcal{C} \subseteq \mathbb{F}_q^{m\times n}$, the *adjoint code* of $\mathcal{C}$ is

$$\mathcal{C}^\top = \left\{ C^\top : C \in \mathcal{C} \right\}.$$

It is clear that, for an $(m,n,q;d)$-code $\mathcal{C}$, the parameters of $\mathcal{C}^\top$ are $(n,m,q;d)$. Therefore, whenever $\mathcal{C}$ and $\mathcal{D}^\top$, for $\mathcal{D}$ as in Lemma 3.4.21, are not equivalent, we will say that $\mathcal{C}$ is not equivalent to a punctured generalised twisted Gabidulin code.

Theorem 3.4.23 ([104, Theorem 6.3]). *Let r, m, h be positive integers such that $h+1$ divides rm, $m \geq h+3$, and $(m,h) \neq (4,1)$. Let $\mathcal{C} = \mathcal{C}_{U,G}$ be the MRD code with parameters $\left(m, \frac{rm}{h+1}, q; m-h\right)$ defined in Theorem 3.4.3. If $h+1$ does not divide r, then $\mathcal{C}$ is not equivalent to any punctured generalised Gabidulin code nor to any punctured generalised twisted Gabidulin code.*

Proof. Let $\rho\colon \mathbb{F}_{q^m} \to \mathbb{F}_{q^n}$ be an $\mathbb{F}_q$-linear map of rank $n \leq m$. Suppose that $\mathcal{C}$ is equivalent to $\mathcal{D}^\top$, where $\mathcal{D}$ is either $\rho \circ \mathcal{G}_{k,\sigma}$ or $\rho \circ \mathcal{H}_{k,\sigma}(\eta, c)$. By equating the minimum distance of $\mathcal{C}$ to $\mathcal{D}^\top$, we get

$$m - h = m - k + 1 \quad \Longrightarrow \quad k = h + 1.$$

Since $m > h + 2$ and $(m, h) \neq (4, 1)$, we can apply Lemma 3.4.21 with $N = m$ and $M = \frac{rm}{h+1}$ in combination with the fact that $L(\mathcal{C}^\top) = R(\mathcal{C})^\top$ and $R(\mathcal{C}^\top) = L(\mathcal{C})^\top$; see [66, Proposition 4.2]. Hence, $|L(\mathcal{D}^\top)| = q^\ell$, where ℓ divides $\frac{rm}{h+1}$. As $\mathcal{C}$ and $\mathcal{D}^\top$ are equivalent, by Theorem 3.1.12(ii) we have $|L(\mathcal{C})| = |L(\mathcal{D}^\top)| = q^\ell$. Then $m = \ell$, whence $m \mid \frac{rm}{h+1}$, which contradicts $(h + 1) \nmid r$. ∎

The authors of [104] conclude their discussion with examples demonstrating that truly new MRD codes, inequivalent to the (at that point) known families, can indeed be constructed via this geometric framework. Crucially, this equivalence shifts the central challenge from an algebraic coding problem to a geometric one: the existence and construction of subspaces U with the required h-scattered property. We will cover specific constructions of such subspaces in Section 3.5.

3.4.4 Intersection numbers obtained via MRD codes

So far, we have used h-scattered subspaces primarily to construct $\mathbb{F}_q$-linear MRD codes, which were shown to be inequivalent to AGTG codes under a single condition. However, the correspondence between these objects also improves our understanding of the underlying geometry. In particular, we will be able to determine the intersection numbers of a linear set, i.e. the possible sizes of the intersection between the linear set and hyperplanes, as well as the number of hyperplanes realising each size.

In this analysis, we follow [104], which uses the methodology first established by Sheekey and Van de Voorde [96]. Their work connects the rank distribution of a code with the intersection numbers of the corresponding linear set, specifically for the case $h = r - 1$, which they call "scattered linear sets with respect to hyperplanes".

First of all, we define linear sets and the weight of points and subspaces with respect to them. These concepts allow us to formally define h-scattered linear sets. A more detailed treatment of linear sets, including some geometric properties, can be found in [102].

Definition 3.4.24 (Linear set). A set L of points in $\Omega = \mathrm{PG}(r-1, q^m)$ is called an *$\mathbb{F}_q$-linear set of rank s* if there exists an s-dimensional $\mathbb{F}_q$-subspace U of $V = V(r, q^m)$, such that $L = L_U$, where

$$L_U := \left\{ \langle \mathbf{u} \rangle_{\mathbb{F}_{q^m}} : \mathbf{u} \in U \setminus \{\mathbf{0}\} \right\}.$$

Here, the notation $\langle \mathbf{u} \rangle_{\mathbb{F}_{q^m}}$ reflects the fact that all $\mathbb{F}_{q^m}$-multiples of $\mathbf{u}$ define the same projective point in Ω.

Definition 3.4.25 (Weight). Let $V = V(r, q^m)$ and let U be an $\mathbb{F}_q$-subspace of V. Let L_U be an $\mathbb{F}_q$-linear set in $\Omega = \mathrm{PG}(V, \mathbb{F}_{q^m})$ defined by the subspace U. The *weight* of a point $P = \langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}} \in \Omega$ in L_U is defined as the dimension of the intersection of U and the one-dimensional $\mathbb{F}_{q^m}$-subspace corresponding to P:

$$w_{L_U}(P) := \dim_{\mathbb{F}_q}(U \cap \langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}}).$$

For an $(k-1)$-dimensional subspace $\pi = \mathrm{PG}(W, \mathbb{F}_{q^m})$, with W a k-dimensional $\mathbb{F}_{q^m}$-subspace of V, we define the *weight* of π with respect to L_U by

$$w_{L_U}(\pi) := \dim_{\mathbb{F}_q}(U \cap W).$$

Definition 3.4.26 (h-scattered linear set). If U is a (maximum) h-scattered $\mathbb{F}_q$-subspace of $V = V(r, q^m)$, L_U is said to be a (maximum) *h-scattered $\mathbb{F}_q$-linear set* in $\Omega = \mathrm{PG}(V, \mathbb{F}_{q^m})$. In other words, an $\mathbb{F}_q$-linear set L_U of Ω is h-scattered if

- $\langle L_U \rangle_{\mathbb{F}_{q^m}} = \Omega$; and
- for every $(h-1)$-subspace $\mathcal{S} = \mathrm{PG}(S, \mathbb{F}_{q^m})$ of Ω, we have $w_{L_U}(\mathcal{S}) \leq h$.

In this context, Lemma 3.4.15 is applicable and if L_U is an h-scattered $\mathbb{F}_q$-linear set of rank $\frac{rm}{h+1}$ in Ω, then for every hyperplane $\mathcal{H}$ of Ω we have

$$\frac{rm}{h+1} - m \leq w_{L_U}(\mathcal{H}) \leq \frac{rm}{h+1} - m + h.$$

The following theorem gives the weight distribution, that is, how many hyperplanes of Ω have a certain weight, in this possible interval.

Theorem 3.4.27 ([104, Theorem 7.1]). *Let L_U be an h-scattered $\mathbb{F}_q$-linear set of rank $\frac{rm}{h+1}$ in $\Omega = \mathrm{PG}(r-1, q^m)$. For every $i \in \{0, \dots, h\}$, the number of hyperplanes of weight $\frac{rm}{h+1} - m + i$ in L_U is given by*

$$t_i = \frac{1}{q^m - 1} \begin{bmatrix} m \\ i \end{bmatrix}_q \sum_{j=0}^{h-i} (-1)^j \begin{bmatrix} m-i \\ j \end{bmatrix}_q q^{\binom{j}{2}} \left(q^{\frac{rm(h-i-j+1)}{h+1}} - 1 \right). \tag{3.9}$$

In particular, $t_i > 0$ for every $i \in \{0, \dots, h\}$.

Proof. Let $\sigma, \sigma', \perp, \perp'$ be defined as in Section 3.4.2, and $\mathcal{H} = \mathrm{PG}(H, \mathbb{F}_{q^m})$ be a hyperplane of Ω with weight $\frac{rm}{h+1} - m + i$ w.r.t. L_U. By Proposition 3.4.11(iii) and (iv), we get

$$\dim_{\mathbb{F}_q}(U^{\perp'} \cap H^{\perp}) = \dim_{\mathbb{F}_q}(U \cap H) + rm - \frac{rm}{h+1} - (r-1)m = i \leq h,$$

whence the weight in $L_{U^{\perp'}}$ of a point (dual of hyperplane) of Ω is at most h. By ordinary duality, the $\mathbb{F}_q$-linear set $L_{U^{\perp'}}$ has rank $\frac{hrm}{h+1}$. Furthermore, the number of points of Ω with weight i in $L_{U^{\perp'}}$ is equal to the number t_i of hyperplanes with weight $\frac{rm}{h+1} - m + i$ in L_U, for every $i \in \{0, \dots, h\}$. Let $W = V(\frac{rm}{h+1}, q)$ and $G \colon V \to W$ be an $\mathbb{F}_q$-linear map with $\ker(G) = U^{\perp'}$. By Theorem 3.4.3, the code $\mathcal{C}_{U^{\perp'},G} = \{ \Gamma_{\mathbf{v}} : \mathbf{v} \in V \}$ is an MRD code.

Note that, for every $i \in \{0, \dots, h\}$, t_i is equal to the number of maps in $\mathcal{C}_{U^{\perp'},G}$ having rank $m-i$, divided by $q^m - 1$. Indeed, consider a point $P = \langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}}$ of weight i in $L_{U^{\perp'}}$. By definition, its weight is $i = \dim_{\mathbb{F}_q}(\langle \mathbf{v} \rangle_{\mathbb{F}_{q^m}} \cap U^{\perp'})$. Since $\ker(G) = U^{\perp'}$, this intersection is precisely the kernel of the map $\Gamma_{\mathbf{v}}$, so $\dim_{\mathbb{F}_q}(\ker(\Gamma_{\mathbf{v}})) = i$. As the domain of $\Gamma_{\mathbf{v}}$ is $\mathbb{F}_{q^m}$ (which has dimension m over $\mathbb{F}_q$), the rank-nullity theorem implies that $\mathrm{rank}(\Gamma_{\mathbf{v}}) = m - i$. Finally, as every projective point P corresponds to $q^m - 1$ non-zero scalar multiples $\lambda \mathbf{v}$, each yielding a map $\Gamma_{\lambda \mathbf{v}}$ of the same rank, we have $t_i = A_{m-i}/(q^m - 1)$, where A_{m-i} denotes the number of codewords in $\mathcal{C}_{U^{\perp'},G}$ of rank $m - i$.

By Theorem 3.4.3, Equation (3.9) follows. Furthermore, Lemma 3.4.2 implies $t_i > 0$ because $\mathcal{C}_{U^{\perp'},G}$ is an MRD code. ∎

Corollary 3.4.28 ([104, Corollary 7.2]). *Let L_U be an h-scattered $\mathbb{F}_q$-linear set of rank $\frac{rm}{h+1}$ in $\Omega = \mathrm{PG}(r-1, q^m)$. For every hyperplane $\mathcal{H}$ of Ω, we have*

$$|\mathcal{H} \cap L_U| \in \left\{ \theta_{\frac{rm}{h+1} - m - 1}, \dots, \theta_{\frac{rm}{h+1} - m + h - 1} \right\},$$

where $\theta_{k-1} = \frac{q^k - 1}{q - 1}$ for all k. Furthermore, for every $i \in \{0, \dots, h\}$, the number of hyperplanes $\mathcal{H}$ of Ω satisfying $|\mathcal{H} \cap L_U| = \theta_{\frac{rm}{h+1} - m + i - 1}$ is t_i, as in Equation (3.9).

In summary, these results demonstrate that the geometry of a maximum h-scattered linear set, specifically its interaction with hyperplanes, is controlled by the parameters of the associated MRD code.

3.4.5 Further generalisations and geometric connections

The geometric results derived above regarding h-scattered subspaces have recently been placed into a broader theoretical framework by Marino, Neri and Trombetti [67]. They used the concept of (s,k)-evasive subspaces – first introduced for general s in [19] – which generalises the notion of h-scattered subspaces. While an h-scattered subspace is defined by intersecting h-dimensional subspaces with a bound equal to that dimension, an (s,k)-evasive subspace decouples these parameters, fixing the dimension k of the testing subspaces while allowing the intersection bound s to vary independently.

Definition 3.4.29 (Evasive subspace). Let U be an $\mathbb{F}_q$-subspace of $V = V(r,q^m)$. For integers s,k with $s \geq 0$ and $1 \leq k \leq r$, the subspace U is called (s,k)-evasive if for every k-dimensional $\mathbb{F}_{q^m}$-subspace H of V, we have

$$\dim_{\mathbb{F}_q}(U \cap H) \leq s.$$

Remark 3.4.30. An h-scattered subspace corresponds to the specific case of an (h,h)-evasive subspace.

The authors of [67] establish a precise correspondence between the evasiveness of a certain subspace and the generalised rank weights of an associated (vector) rank-metric code (see Definition 3.2.5 and Notation 3.2.6). This requires introducing a specific geometric object, the q-system.

Definition 3.4.31 (q-system). An $[n,k,d]_{q^m/q}$*-system* $\mathcal{U}$ is an n-dimensional $\mathbb{F}_q$-subspace of $\mathbb{F}_{q^m}^k$ such that $\langle \mathcal{U} \rangle_{\mathbb{F}_{q^m}} = \mathbb{F}_{q^m}^k$. The integer d is defined as

$$d := n - \max\{\dim_{\mathbb{F}_q}(H \cap \mathcal{U}) : H \text{ is an } \mathbb{F}_{q^m}\text{-hyperplane of } \mathbb{F}_{q^m}^k\}.$$

For brevity, when the parameters are irrelevant, we will simply say that $\mathcal{U}$ is a *q-system.*

Definition 3.4.32 (Generalised rank weight). Let $\mathcal{C}$ be an $[n,k,d]_{q^m/q}$-code, and let $1 \leq s \leq k$. The s*-th generalised rank weight* of $\mathcal{C}$ is the integer

$$d_{R,s} := \min\{\dim_{\mathbb{F}_{q^m}}(A) : A \in \Lambda_q(n,m), \dim_{\mathbb{F}_{q^m}}(A \cap \mathcal{C}) \geq s\},$$

where $\Lambda_q(n,m)$ denotes the set of Galois closed $\mathbb{F}_{q^m}$-subspaces of $\mathbb{F}_{q^m}^n$, i.e.

$$\theta(A) := \{\, (\theta(a_1), \ldots, \theta(a_n)) : (a_1, \ldots, a_n) \in A \,\} = A,$$

for every $\theta \in \mathrm{Gal}(\mathbb{F}_{q^m}/\mathbb{F}_q)$.

Definition 3.4.33. Let s be a positive integer. An $[n,k,d]_{q^m/q}$-code $\mathcal{C}$ is s*-MRD* if $d_{R,s}(\mathcal{C}) = n-k+s$.

Remark 3.4.34. While every 1-MRD code is MRD, the converse holds only when $n \leq m$. Indeed, the 1-MRD property requires the first generalised rank weight (the minimum distance d) to equal $n-k+1$. However, the definition of an MRD code is based on attaining the Singleton bound (3.1). When $n > m$, this bound evaluates to $mk \leq n(m-d+1)$, which rearranges to $d \leq m - \left\lceil \frac{km}{n} \right\rceil + 1$. Because $n > m$, this theoretical maximum is strictly less than $n-k+1$, making that limit unreachable. An MRD code will attain this optimal distance – satisfying the definition of an MRD code – while failing the condition to be 1-MRD.

It was shown that there is a correspondence between the evasiveness of a q-system and the generalised weights of its associated code.

Lemma 3.4.35 ([87, Theorem 2]). *We can associate an* $[n,k,d]_{q^m/q}$*-system* $\mathcal{U}$ *to every* $[n,k,d]_{q^m/q}$*-code* $\mathcal{C}$*, where codewords of* $\mathcal{C}$ *of rank weight* w *correspond to* $\mathbb{F}_{q^m}$*-hyperplanes* H *of* $\mathbb{F}_{q^m}^k$ *with* $\dim_{\mathbb{F}_q}(H \cap \mathcal{U}) = n - w$.

Theorem 3.4.36 ([67, Theorem 3.3]). *Let $\mathcal{C}$ be a non-degenerate $[n,k,d]_{q^m/q}$-code, and let $\mathcal{U}$ be its associated $[n,k,d]_{q^m/q}$-system. Then, the following are equivalent:*

(i) $\mathcal{U}$ is (h,r)-evasive,

(ii) $d_{R,k-h}(\mathcal{C}) \geq n-r$,

(iii) $d_{R,r-h+1}(\mathcal{C}^\perp) \geq r+2$, where $\mathcal{C}^\perp = \left\{ \mathbf{v} \in \mathbb{F}_{q^m}^n : c\mathbf{v}^\top = 0, \forall c \in \mathcal{C} \right\}$.

The theorem above allows one to recover the result of Theorem 3.4.3, yet it imposes a noteworthy restriction. While Zini and Zullo [104] consider $\mathbb{F}_q$-linear codes, Marino et al. [67] focus on vector rank-metric codes that correspond to $\mathbb{F}_{q^m}$-linear subspaces. This distinction is important for cryptographic applications: we would like codes that are not $\mathbb{F}_{q^m}$-linear to be less susceptible to structural attacks.

Beyond generalising previous results, this extended framework offers tools for constructing and classifying codes with targeted properties. The authors leverage this connection to characterise near-MRD codes, i.e. codes that miss the Singleton bound by one but retain optimal generalised weights for higher parameters. Through the geometric lens of q-systems, they establish strict bounds on the length of such codes. These bounds are particularly relevant for identifying candidates for secure random network coding, as a code's generalised weights directly quantify its ability to minimise information leakage to a wiretapper. Furthermore, the theory extends to the class of so-called quasi-MRD codes.

Finally, it is worth noting that these q-systems also yield minimal codes under specific conditions, due to a connection with linear cutting blocking sets [1]. While minimal codes are significant for secret sharing schemes, we will not explore them further, as this thesis focuses on MRD codes for encryption. Nevertheless, it is evident that geometric connections as described in this section are a driving force in the cryptographic study of MRD codes.

3.5 Constructing maximum h-scattered subspaces

The correspondence established in the previous section effectively shifts the task of constructing $\mathbb{F}_q$-linear MRD codes from an algebraic setting to the geometric problem of constructing maximum h-scattered subspaces. Notably, Sheekey's seminal 2016 work [94] pioneered this perspective, and he offered an explicit construction to derive twisted Gabidulin codes. As previously discussed, this correspondence was later generalised and expanded by Zini and Zullo [104]. Nevertheless, determining the existence of and constructing such h-scattered subspaces for the full range of parameters r, m, h remains an important open problem.

In this section, we briefly mention some methods employed for constructing maximum h-scattered subspaces. Interestingly, this geometric problem is frequently translated back into the algebraic domain of linearised polynomials, allowing researchers to leverage the tools introduced in Section 3.2.3. We focus primarily on the case $r = 2$ (scattered with respect to lines), where the theory is most developed, before extending the discussion to higher dimensions.

3.5.1 Constructions on the projective line

We will first focus on maximum h-scattered $\mathbb{F}_q$-subspaces of $V = V(2, q^m)$. Note that h necessarily equals 1 in this setting, and a maximum 1-scattered (or simply scattered) subspace has dimension $\frac{rm}{h+1} = \frac{2m}{2} = m$.

Let U be such a maximum scattered subspace. Then we can associate a linear set L_U of rank m with it on the projective line $\mathrm{PG}(1, q^m)$. Since the group $\mathrm{PGL}_2(\mathbb{F}_{q^m})$ acts 3-transitively on the points of $\mathrm{PG}(1, q^m)$, and the size of L_U is strictly less than the size of the projective line, we can always find a coordinate transformation such that the point at infinity $\langle (0,1) \rangle_{\mathbb{F}_{q^m}}$ does not belong to L_U. Consequently, the underlying subspace U has a trivial intersection with the subspace $\langle (0,1) \rangle_{\mathbb{F}_{q^m}}$,

meaning U can be uniquely represented as the graph of an $\mathbb{F}_q$-linear map $f : \mathbb{F}_{q^m} \to \mathbb{F}_{q^m}$, i.e. the set of vectors $\{ (x, f(x)) : x \in \mathbb{F}_{q^m} \}$. Thus, we can write

$$U = U_f = \{ (x, f(x)) : x \in \mathbb{F}_{q^m} \},$$

where f is a linearised polynomial.

This representation leads to the following definition, where the ratio $f(x)/x$ is interpreted as the slope of the vector $(x, f(x))$ in the graph U_f.

Definition 3.5.1 (Scattered polynomial). A linearised polynomial $f \in \widetilde{\mathcal{L}}_{m,q}[x]$ is called a *scattered polynomial* if, for any $x, y \in \mathbb{F}_{q^m} \setminus \{0\}$, the condition

$$\frac{f(x)}{x} = \frac{f(y)}{y}$$

implies that x and y are $\mathbb{F}_q$-linearly dependent.

The geometric property of the subspace U_f being scattered is linked to the algebraic properties of the polynomial f and the parameters of the associated rank-metric code. This connection is formalised in the following theorem.

Theorem 3.5.2 ([94, Section 5]). *Let $f \in \widetilde{\mathcal{L}}_{m,q}[x]$. Then, the following statements are equivalent:*

(i) U_f is a maximum scattered subspace of $V(2, q^m)$;

(ii) f is scattered;

(iii) The set of linearised polynomials

$$\mathcal{C}_f = \langle x, f(x) \rangle_{\mathbb{F}_{q^m}} = \{\alpha x + \beta f(x) : \alpha, \beta \in \mathbb{F}_{q^m}\} \subseteq \widetilde{\mathcal{L}}_{m,q}[x]$$

is a maximum-rank distance (MRD) code with parameters $(m, m, q; m-1)$.

Hence, scattered polynomials f can facilitate the construction of maximum scattered subspaces and thus MRD codes. In the following, we will present some distinct families of scattered polynomials.

Known families and classification. The search for scattered polynomials has provided us with several distinct families, each giving rise to maximum scattered subspaces (and thus MRD codes). The significance of these families lies in their equivalence classes: if two scattered polynomials f and g generate subspaces U_f and U_g that are not equivalent under the action of $\Gamma\mathrm{L}_2(\mathbb{F}_{q^m})$, they yield candidates for inequivalent MRD codes.

The two most basic, yet important, families of scattered polynomials $f \in \widetilde{\mathcal{L}}_{m,q}[x]$ discovered to date are:

- **The pseudoregulus family (monomials):** The simplest example of a scattered polynomial is the monomial

  $$f(x) = x^{q^s},$$

  where $\gcd(s, m) = 1$. The linear set corresponding to the subspace U_f defined by this polynomial forms a so-called pseudoregulus, hence the name. Algebraically, the rank-metric code associated with this family is equivalent to the classical Gabidulin code. This serves as the baseline for classification; any new construction is first checked to ensure it is not merely a coordinate transformation of this family.

- **The Lunardon-Polverino family (binomials):** Introduced in [63] and later generalised, this family consists of binomials of the form

$$f(x) = x^{q^s} + \eta x^{q^{m-s}},$$

where $m \geq 4$, $\gcd(s, m) = 1$ and $N_{q^m/q}(\eta) \notin \{0, 1\}$. Crucially, for appropriate parameters, the subspaces generated by these polynomials are strictly inequivalent to the pseudoregulus type, providing a genuine expansion of the known class of MRD codes.

More families can be found in [59, Section 2.1].

To verify that a newly discovered scattered polynomial yields a truly "new" MRD code, one can rely on geometric invariants, again underlining the fruitful interplay between coding theory and finite geometry.

3.5.2 Higher dimensions

When extending to dimensions $r > 2$, the construction of maximum h-scattered subspaces becomes significantly more involved. The fundamental representation as a simple graph of a polynomial – which was central to the $r = 2$ case – is no longer sufficient for the vast majority of parameters. Specifically, whenever the target dimension $\frac{rm}{h+1}$ exceeds the extension degree m, it is impossible to represent the subspace U using a single variable $x \in \mathbb{F}_{q^m}$. This is because the dimension of a graph $\{ (x, f(x)) \}$ is strictly limited by the dimension of its input domain, which is exactly m.

However, there is a notable exception: the specific case where $h = r - 1$. In this setting, the required dimension of the subspace is exactly $\frac{rm}{(r-1)+1} = m$. Since the dimension matches the extension degree, the graph representation can be naturally generalised. Such a subspace can be uniquely represented as the graph of a vector-valued map, defined by a set of polynomials:

$$U_{f_1,\dots,f_{r-1}} = \{ (x, f_1(x), \dots, f_{r-1}(x)) : x \in \mathbb{F}_{q^m} \}, \tag{3.10}$$

where $f_1, \dots, f_{r-1} \in \widetilde{\mathcal{L}}_{m,q}[x]$. However, for the subspace U in (3.10) to be maximum $(r-1)$-scattered, the defining sequence $f_1, \dots, f_{r-1}$ must satisfy specific conditions; see, for example, [3].

For all other cases where the target dimension exceeds m, alternative construction methods are required. One effective approach is a direct sum construction, as established in [4], where multiple copies of 2-dimensional subspaces U_f serve as fundamental building blocks. By combining these, one can form maximum 1-scattered linear sets in higher-dimensional projective spaces $PG(r-1, q^m)$ for even r.

Constructing MRD codes via scattered subspaces remains an active area of research, with recent classification results that guide the search towards effective constructions.

4 Cryptanalysis of GPT cryptosystems

Establishing the security of cryptographic primitives is a complex challenge. Indeed, only the one-time pad – which encrypts each plaintext with a unique, random secret key of equal length – achieves perfect secrecy [93]. Consequently, modern cryptography relies on two alternative foundations for assuming security: either the cryptographic primitive is mathematically reduced to a problem known to be NP-complete (i.e. hard to solve), or it has withstood extensive cryptanalytic research over decades resisting all attempted attacks.

The RSA algorithm serves as an excellent example of this latter category. While its security is widely assumed to rest on the hardness of integer factorisation, this foundation is theoretically shaky for two reasons. First of all, factorisation is not known to be NP-complete. Secondly, it has not been proven that factorisation is the only path to breaking RSA; an adversary could theoretically decrypt messages without ever factoring the modulus. Therefore, the assumption that RSA is secure is empirical rather than provable: it is accepted because decades of extensive research have failed to discover a polynomial-time factorisation algorithm or any alternative method to break the cryptographic primitive.

In contrast, the GPT cryptosystem appears, at first glance, to rest on firmer theoretical foundations. As established in Chapter 2, the general decoding problem of a random linear code in the rank metric is believed to be NP-hard (supported by a probabilistic reduction from SDP). However, practical public-key cryptography cannot rely on purely random codes; efficient decoding (and thus decryption) requires the use of codes with a strong algebraic structure, such as the Gabidulin codes. While the decoding problem is computationally intractable for random instances, the specific structure required for efficient usage exposes the scheme to structural attacks.

Generic decoding algorithms are typically ciphertext-only attacks: they attempt to recover the plaintext $\mathbf{x}$ or the error vector $\mathbf{e}$ from a specific ciphertext, leaving the system intact for future use. Structural attacks, conversely, are key-recovery attacks. By exploiting the algebraic structure of the code, an adversary seeks to reconstruct an equivalent private key, thereby compromising the entire cryptosystem and allowing the decryption of any past or future messages.

While this chapter primarily focuses on the structural cryptanalysis of the GPT cryptosystem, we begin by briefly addressing non-structural attacks, showing why smaller public key sizes (as compared to McEliece) are possible. Subsequently, we review Gibson's early attack, which first exposed weaknesses in the scheme, before presenting Overbeck's attack, which effectively broke the original implementation based on Gabidulin codes for all parameter sets. Finally, we examine extensions of Overbeck's attack, demonstrating how subsequent attempts to mask the algebraic structure of the used Gabidulin code were eventually broken too. This analysis illustrates the perpetual cat-and-mouse game between cryptographers and cryptanalysts.

Representation of rank-metric codes. For the GPT cryptosystem, the vector representation of rank-metric codes, as detailed in Section 3.2.2, is used. To facilitate efficient encryption and decryption, the GPT cryptosystem specifically employs codes from the Gabidulin family. Rather than defining them via a general subspace as before, we focus on the specific representation used in the cryptosystem's key generation. A Gabidulin code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ of dimension k is defined by a generator vector $\mathbf{g} = (g_1, \ldots, g_n) \in \mathbb{F}_{q^m}^n$, where the components $g_1, \ldots, g_n$ are linearly independent over the base field $\mathbb{F}_q$. In that case, we say that $\mathcal{C}$ is an $[n, k]$ Gabidulin code. The generator matrix $G \in \mathbb{F}_{q^m}^{k \times n}$ is constructed by applying the Frobenius automorphism coordinate-wise to this vector. Specifically, the i-th row of G is given by the $(i-1)$-th Frobenius power of $\mathbf{g}$, denoted as $\mathbf{g}^{[i-1]} = (g_1^{q^{i-1}}, \ldots, g_n^{q^{i-1}})$. Thus, the

generator matrix G of an $[n, k]$ Gabidulin code takes the form

$$G = \begin{pmatrix} g_1 & g_2 & \cdots & g_n \\ g_1^{[1]} & g_2^{[1]} & \cdots & g_n^{[1]} \\ \vdots & \vdots & \ddots & \vdots \\ g_1^{[k-1]} & g_2^{[k-1]} & \cdots & g_n^{[k-1]} \end{pmatrix}. \tag{4.1}$$

This distinct algebraic structure allows for efficient decoding but, as we will see, also introduces the vulnerabilities exploited by multiple attacks.

For the remainder of this chapter, we will simply use the term "rank" to refer to the (vector or matrix) rank over the base field $\mathbb{F}_q$ (see Definition 3.2.5). However, because structural cryptanalysis frequently manipulates matrices defined over the extension field $\mathbb{F}_{q^m}$, ambiguity can arise. Whenever clarity is required to distinguish between the rank over the base field and the extension field, we will explicitly rely on the notation $\mathrm{rk}_q(\cdot)$ established in Notation 2.5.1.

4.1 Non-structural attacks

As discussed in Section 2.4.1, non-structural attacks (or generic decoding attacks) disregard the specific algebraic structure of the code used in the cryptosystem. Instead, they treat the public generator matrix $\hat{G}$ as if it were chosen uniformly at random. Therefore, the security of the GPT cryptosystem against this class of attacks relies entirely on the hardness of the underlying rank syndrome decoding (RSD) problem.

The consensus within the cryptographic community remains that the RSD problem is significantly harder than the equivalent problem in the Hamming metric. This claim was central to the original proposal by Gabidulin, Paramonov, and Tretjakov in 1991 [29]. The argument rests on the fundamental difference in the search space of the error vector. In the binary Hamming metric, an error of weight t corresponds to a set of t non-zero coordinates. A generic attack, such as information set decoding, succeeds by correctly guessing error-free positions. The complexity is roughly given by the binomial coefficient $\binom{n}{t}$, which represents the number of possible error vectors of weight t.

In the rank metric, however, an error of rank weight t corresponds to a t-dimensional subspace within the ambient vector space $\mathbb{F}_{q^m}^n$. The number of such subspaces is given by the Gaussian binomial coefficient (see Definition 3.1.6), meaning that the search space for rank errors scales with the field size q^m. Because rank-metric codes typically operate over a large extension field, generic rank decoding is orders of magnitude harder than generic Hamming decoding for comparable parameters. This inherent hardness allows rank-metric cryptosystems to employ much smaller keys while maintaining the same security level against non-structural attacks.

Over the last three decades, cryptanalysts have sought to improve upon this exhaustive search. The most effective solvers to date model the decoding problem as a system of equations. However, for parameters of cryptographic interest, these algebraic systems are typically underdetermined; the number of variables exceeds the number of independent equations, rendering a direct solution impossible.

To overcome this problem, hybrid strategies are employed that combine an initial guess with algebraic solving. An attacker first guesses partial information about the introduced error, which reduces the number of variables, thereby transforming the system into an overdetermined one that can be solved. A significant recent advancement in this direction for RSD is the hybrid guessing technique introduced in [5].

When the state-of-the-art regarding non-structural attacks advances, more secure parameters (e.g. error vectors with larger rank) are required. However, the exponential complexity is retained and hence the hardness assumption of the generic rank decoding problem remains widely believed to hold. Unfortunately, the vulnerability of the GPT cryptosystem does not lie in the general hardness of rank decoding, but rather in the specific, structured nature of the Gabidulin codes used to instantiate it. Therefore, the structural attacks pose the greatest threat to the GPT cryptosystem.

4.2 Gibson's attacks

The first serious cryptanalytic blow to the GPT cryptosystem, as presented in Section 2.5, was dealt by Gibson in 1995 [34]. His work highlighted that while the underlying rank decoding problem might be hard, the specific algebraic structure of the masking matrices – specifically the distortion matrix X – can be exploited if the cryptosystem's parameters are not chosen with care.

4.2.1 The rank-1 attack

Recall the structure of the public key in the original GPT cryptosystem: $\hat{G} = S(G + X)$, with scrambler matrix S and distortion matrix X. In the initial proposal [29], the authors constructed the distortion matrix X such that $s = \mathrm{rk}_{q^m}(X) = 1$. In particular, the distortion matrix decomposes as $X = \mathbf{r}^\top \mathbf{c}$, where $\mathbf{r} \in \mathbb{F}_{q^m}^k$ acts as row distortion vector and $\mathbf{c} \in \mathbb{F}_{q^m}^n$ as column distortion vector. Therefore, the public key can be rewritten as:

$$\hat{G} = SG + (S\mathbf{r}^\top)\mathbf{c}.$$

Gibson [34] showed that by manipulating the rows of $\hat{G}$, one can construct a system of non-linear equations over $\mathbb{F}_{q^m}$ involving the unknown components $\mathbf{g}$ (generator vector for G), $\mathbf{r}$, and $\mathbf{c}$. Notably, this analysis effectively eliminates the scrambler matrix S from the set of secret variables. Gibson demonstrated that S acts merely as a linear transformation linking the rows of the public matrix $\hat{G}$ to the private code structure; once the parameters $\mathbf{g}$, $\mathbf{r}$, and $\mathbf{c}$ are recovered, S is uniquely determined and can be efficiently computed from the public key.

In short, to solve the system of non-linear equations, write the public key $\hat{G}$ in systematic form as $\hat{G} = (I_k \mid Y)$, where I_k denotes the $k \times k$ identity matrix. Accordingly, split the column distortion vector into $\mathbf{c} = (\mathbf{u} \mid \mathbf{v})$, where $\mathbf{u}$ contains the first k coordinates and $\mathbf{v}$ contains the remaining $n - k$ coordinates. By guessing a single component f_2 of the vector $\mathbf{f} = (f_1, f_2, \ldots, f_{n-k}) = \mathbf{u}Y + \mathbf{v}$, under the in practice realistic assumption that $n - k \geq 3$, the system becomes solvable. For the parameters $q = 2$, $m = n = 20$, $k = 12$, and $\hat{t} = 3$ originally considered in [29] as numerical example, the search space for guessing f_2 is approximately $q^m = 2^{20}$, which is small enough to be exhausted in feasible time.

It is important to note that the recovered generator vector, and scrambler and distortion matrices may differ from the original private key. However, they combine to form the same public key, and crucially, the recovered generator matrix allows for the correct decryption of the ciphertext.

The overall complexity of the algorithm (adapted for general s in [27]) is

$$\mathcal{O}\left(m^3(n-k)^3 q^{ms}\right).$$

This attack effectively broke the $s = 1$ instance of the GPT cryptosystem. While increasing the code size n can prevent the attack, the primary advantage of the Gabidulin version, namely its compact public key compared to the McEliece system, is then largely lost.

4.2.2 The general rank attack

In response to Gibson's first attack Gabidulin proposed increasing the rank of the distortion matrix X over the extension field $\mathbb{F}_{q^m}$ from $s = 1$ to $s = \ell$ (where $\ell = \mathrm{rk}_q(X)$) in [27, Section 5.3]. The idea was that such a distortion would increase the number of possibilities for the guessed variable, rendering the exhaustive search computationally infeasible.

However, Gibson presented a new attack in 1996 [35] that does not rely on solving the non-linear system of equations directly. Instead of leveraging the algebraic relationship between the public and private keys, Gibson focused on the structural decomposition of the distortion matrix X. He observed that X can be factored into a row component A over $\mathbb{F}_{q^m}$ and a column component B over $\mathbb{F}_q$.

Gibson introduced a parameter d, called the *deficiency* of the public key, which measures how much information the public matrix leaks about this hidden column component B. The attack demonstrates

that the computational effort to recover B is exponential in d. Crucially, for the specific parameter $s = \ell$ Gabidulin chose to be secure against the first attack, the deficiency is extremely low: often $d = 0$ or $d = 1$.

Furthermore, because the column distortion component B is restricted to the base field $\mathbb{F}_q$ rather than the extension field $\mathbb{F}_{q^m}$, the number of possible candidates for its entries is exponentially smaller. This allows the attacker to recover B iteratively, column by column, by solving systems of linear equations over $\mathbb{F}_q$.

Once B is known, the attack is essentially complete. The attacker can compute a matrix H such that $BH = 0$. By right-multiplying the public key $\hat{G}$ by H, the distortion matrix is completely removed:

$$\hat{G}H = S(G + AB)H = SGH + SABH = SGH. \tag{4.2}$$

The attacker can then find a valid generator matrix G' and a corresponding scrambler matrix S'. While these recovered matrices may differ from the original private key, they share the exact same error-correcting capability as the original, allowing decryption of any ciphertext.

The complexity of this general rank attack is given by [80]:

$$\mathcal{O}\left(k^3 + (k+\ell)d \cdot q^{d(k+2)} + (m-k)\ell \cdot q^d\right),$$

where d corresponds to the deficiency parameter. Notably, if $d = 0$, the attack runs in polynomial time. It is also important to highlight that the success of both the rank-1 and the general rank attack relies on some (weak) assumptions regarding the cryptosystem's matrices and parameters. However, it was shown that these assumptions are fulfilled with high probability for random choices of the generator matrix G, the scrambler matrix S, and the distortion matrix X.

Therefore, the attempt to fix the rank-1 vulnerability by setting $s = \ell$ made the system vulnerable to this new attack. To resist Gibson's general rank attack, the system parameters must be carefully chosen to guarantee a high deficiency d, which (again) leads to large(r) public key sizes.

4.2.3 The GGPT cryptosystem

In an attempt to fix the security of the scheme and resist Gibson's second attack, Ourivski and Gabidulin proposed yet another variant in 2003 [78]. This variant masks the public key on both sides and uses two distortion matrices:

$$\hat{G} = S(X_1 \mid G + X_2)T, \tag{4.3}$$

where $X_1 \in \mathbb{F}_{q^m}^{k\times\ell}$ and $X_2 \in \mathbb{F}_{q^m}^{k\times n}$ are random matrices such that $\mathrm{rk}_q(X_2) =: t_2 < t = \lfloor \frac{n-k}{2} \rfloor$, and $T \in \mathrm{GL}_{n+\ell}(\mathbb{F}_q)$. The column scrambler T is used to mix the (distorted) Gabidulin columns with the random columns of X_1. The rationale was that any reduction as in Equation (4.2) using Gibson's second attack would be impossible, obtaining a matrix of the form $SGH + X^*$ instead, where X^* is a residual distortion with residual rank t^*.

It was proven in [78] that the cryptosystem's parameters can be chosen such that the residual rank satisfies $t^* > 2t - t_2$. Since the injected error during encryption has rank $t - t_2$, satisfying this bound ensures that the combined rank of the residual distortion and the injected error strictly exceeds the maximum error-correcting capability t of the underlying (private) Gabidulin code. Therefore, the attacker is faced with an uncorrectable error. By ensuring this condition is met, Ourivski and Gabidulin concluded that Gibson's structural attack is impossible to execute.

However, in 2005, Overbeck [80] showed that this countermeasure is insufficient. He proved that a public key in the form of Equation (4.3) is equivalent to a standard GPT instance (without T and with a single distortion matrix) defined over a larger field and with a larger code length. Specifically, he showed that by viewing the system over this larger field, the scrambled public key takes the form $\hat{G}' = S'(G' + X')$.

Furthermore, Overbeck revealed a second vulnerability that enables an attacker to correct more errors than expected. In the fully-distorted cryptosystem (4.3) from [78], the legitimate receiver must correct both the distortion matrix X_2 (of rank t_2) and the error injected during encryption. This forces the sender to keep the injected error $\hat{t}$ small: $\hat{t} \leq t - t_2$. Overbeck proved that an attacker can compute a transformation of $\hat{G}'$ over the base field $\mathbb{F}_q$ that isolates all the distortion into exactly t_2 columns. By grouping and discarding these specific columns, the attacker obtains a slightly shorter but undistorted Gabidulin code of length $n - t_2$. While this degraded code has a reduced error-correcting capability $\bar{t} = \lfloor \frac{n-t_2-k}{2} \rfloor$, it remains capable of correcting the injected error, since $\bar{t} \geq t - \frac{t_2}{2} > \hat{t}$. Therefore, the attacker can decode any ciphertext.

Despite this successful cryptanalysis, Overbeck did not abandon the concept of column scrambling. In the same work [80], he proposed a generalised GPT (GGPT) cryptosystem designed specifically to resist these structural weaknesses. The novelty of the GGPT cryptosystem lies in its altered decryption mechanism: instead of attempting to correct the distortion, the legitimate receiver discards the distorted columns. This decoupling of the distortion from the error-correction process allows the public error-correcting capability $\hat{t}$ to be chosen significantly larger, circumventing the identified flaw. Therefore, even if an attacker attempts to isolate the distortion using Overbeck's reduction, the injected error $\hat{t}$ is designed to strictly exceed the decoding capability $\bar{t}$ of any degraded code, thereby preventing a successful attack.

The GGPT cryptosystem is defined as follows, using $M_{\{i_1,\ldots,i_k\}}$ to denote the matrix M restricted to the columns $i_1, \ldots, i_k$:

- **System parameters:** $q, n, m, k, \ell, s \in \mathbb{N}$, where $n \leq m$, $\ell < n - k - 1$, and $s \leq \min\{k, \ell\}$.
- **Key generation:** First, generate the following matrices over $\mathbb{F}_{q^m}$:
  - G: $k \times n$ generator matrix of an $[n, k]$ Gabidulin code $\mathcal{G}$ with error-correcting capability t;
  - S: $k \times k$ random non-singular matrix (the row scrambler);
  - X: $k \times \ell$ random matrix with $\mathrm{rk}_{q^m}(X) = s$ and $\mathrm{rk}_q(X) = \ell$ (the distortion matrix).

  Moreover, generate an $n \times n$ random non-singular matrix T over $\mathbb{F}_q$ (the column scrambler). Then, construct the public matrix $\hat{G}$ by distorting only the first ℓ columns of G and scrambling the result:
  $$\hat{G} = S\left(G_{\{1,\ldots,\ell\}} + X \mid G_{\{\ell+1,\ldots,n\}}\right) T.$$
  Choose the public error-correcting capability $1 \leq \hat{t} \leq \frac{n-k-\ell}{2}$. Furthermore, let $\mathcal{D}_{\mathcal{G}}$ be an efficient decoding algorithm for $G_{\{\ell+1,\ldots,n\}}$.
- **Public key:** $(\hat{G}, \hat{t})$
- **Private key:** $(\mathcal{D}_{\mathcal{G}}, S, T)$ or (G, S, T)
- **Encryption:** To encode a plaintext $\mathbf{x} \in \mathbb{F}_{q^m}^k$, choose a vector $\mathbf{e} \in \mathbb{F}_{q^m}^n$ with rank weight $\hat{t}$ at random and compute the ciphertext $\mathbf{c}$ as follows:
  $$\mathbf{c} = \mathbf{x}\hat{G} + \mathbf{e}.$$
- **Decryption:** To decode a ciphertext $\mathbf{c}$, first remove the column scrambler to obtain $\mathbf{c}' = \mathbf{c}T^{-1}$. Since the last $n - \ell$ columns of the public key correspond to an undistorted Gabidulin code, the legitimate receiver discards the first ℓ columns entirely and applies $\mathcal{D}_{\mathcal{G}}$:
  $$\mathbf{x}SG_{\{\ell+1,\ldots,n\}} = \mathcal{D}_{\mathcal{G}}\left(\mathbf{c}'_{\{\ell+1,\ldots,n\}}\right).$$
  From this, we can compute the plaintext $\mathbf{x}$.

As mentioned before, Overbeck showed that all the noise from the distortion matrices in the public key can be isolated. However, to prove this separation in 2005, Overbeck relied on embedding the cryptosystem into a larger field [80]. Later, Otmani et al. [77] proved a similar result, which we will present next.

Remark 4.2.1. For the purpose of the subsequent structural analysis, we may absorb the first ℓ Gabidulin columns into the distortion term and rewrite the public key as $\hat{G} = S\left(X \mid G_{\{\ell+1,\ldots,n\}}\right)T$, where, by abuse of notation, X now denotes the matrix $X := G_{\{1,\ldots,\ell\}} + X$. In this rewritten form, X is an arbitrary matrix in $\mathbb{F}_{q^m}^{k\times\ell}$; it is not assumed (nor required) to retain the rank properties of the original distortion matrix.

To prove that a public key in the form of Equation (4.3) can be reduced to $\hat{G} = S(X^* \mid G^*)T^*$, we first establish a necessary tool and its corollary. The proofs thereof were omitted in the original source.

Proposition 4.2.2 ([77, Proposition 1]). *Let $M \in \mathbb{F}_{q^m}^{k\times n}$ be a matrix with $\mathrm{rk}_q(M) = s < n$. Then, there exist matrices $M^* \in \mathbb{F}_{q^m}^{k\times s}$ with $\mathrm{rk}_q(M^*) = s$ and $T \in \mathrm{GL}_n(\mathbb{F}_q)$ such that*

$$MT = \left(M^* \mid 0_{k\times(n-s)}\right), \tag{4.4}$$

where $0_{k\times(n-s)}$ denotes the $k \times (n-s)$-matrix consisting of all zeros. In particular, for any $\mathbf{x} \in \mathbb{F}_{q^m}^n$ such that $\mathrm{rk}_q(\mathbf{x}) = s$, there exists a matrix $T \in \mathrm{GL}_n(\mathbb{F}_q)$ for which $\mathbf{x}T = \left(\mathbf{x}^* \mid 0_{1\times(n-s)}\right)$, where $\mathbf{x}^* \in \mathbb{F}_{q^m}^s$ and $\mathrm{rk}_q(\mathbf{x}^*) = s$.

Proof. We can select s columns from M that are linearly independent over $\mathbb{F}_q$, forming a basis for an s-dimensional $\mathbb{F}_q$-subspace of $\mathbb{F}_{q^m}^k$. By applying a permutation matrix $P \in \mathrm{GL}_n(\mathbb{F}_q)$, we can reorder the columns of M such that these basis vectors occupy the first s positions: $MP = (M^* \mid M')$, where $M^* \in \mathbb{F}_{q^m}^{k\times s}$ with $\mathrm{rk}_q(M^*) = s$, and M' contains the remaining $n-s$ columns.

Now, every column in M' can be expressed as a linear combination of the columns in M^* using coefficients from $\mathbb{F}_q$. Therefore, there exists a matrix $C \in \mathbb{F}_q^{s\times(n-s)}$ such that $M' = M^*C$.

Using the matrix

$$U = \begin{pmatrix} I_s & -C \\ 0 & I_{n-s} \end{pmatrix} \in \mathrm{GL}_n(\mathbb{F}_q),$$

we find $MPU = (M^* \mid M' - M^*C) = \left(M^* \mid 0_{k\times(n-s)}\right)$. Setting $T = PU \in \mathrm{GL}_n(\mathbb{F}_q)$ proves Equation (4.4).

In particular, the result follows for a vector $\mathbf{x} \in \mathbb{F}_{q^m}^n$ by treating it as a $1 \times n$-matrix. ∎

Corollary 4.2.3 ([77, Corollary 1]). *For any $M \in \mathbb{F}_{q^m}^{k\times n}$ and $\mathbf{m} \in \mathbb{F}_{q^m}^k$, it holds that $\mathrm{rk}_q(\mathbf{m}M) \leq \mathrm{rk}_q(M)$.*

Proof. Let $s = \mathrm{rk}_q(M)$. By Proposition 4.2.2, there exists an invertible matrix $T \in \mathrm{GL}_n(\mathbb{F}_q)$ such that $MT = \left(M^* \mid 0_{k\times(n-s)}\right)$, where $M^* \in \mathbb{F}_{q^m}^{k\times s}$. We get:

$$(\mathbf{m}M)T = \mathbf{m}(MT) = \mathbf{m}\left(M^* \mid 0_{k\times(n-s)}\right) = \left(\mathbf{m}M^* \mid 0_{1\times(n-s)}\right).$$

Because $T \in \mathrm{GL}_n(\mathbb{F}_q)$, we have $\mathrm{rk}_q(\mathbf{m}M) = \mathrm{rk}_q((\mathbf{m}M)T)$.

The coordinates of the transformed vector $\left(\mathbf{m}M^* \mid 0_{1\times(n-s)}\right)$ are entirely contained within its first s positions. Therefore, the dimension of the $\mathbb{F}_q$-vector space spanned by its components can be at most s: $\mathrm{rk}_q(\mathbf{m}M) \leq s = \mathrm{rk}_q(M)$. ∎

Now, we arrive at the result that any full-length distortion matrix applied across the entire Gabidulin code can be isolated into a separate block of columns, leaving an undistorted – but slightly shorter – Gabidulin code exposed.

Proposition 4.2.4 ([77, Proposition 3]). *Let $\hat{G}$ be as in Equation (4.3) and assume $\mathrm{rk}_q(X_2) = t_2$. Then, there exist matrices $T^* \in \mathrm{GL}_{n+\ell}(\mathbb{F}_q)$, $X^* \in \mathbb{F}_{q^m}^{k\times(\ell+t_2)}$, and a matrix G^* that generates an $[n-t_2, k]$ Gabidulin code such that*

$$\hat{G} = S(X^* \mid G^*)T^*.$$

Furthermore, the error-correcting capability t^* of this new Gabidulin code generated by G^* is greater than or equal to the public error-correcting capability $\hat{t}$.

Proof. Since $\mathrm{rk}_q(X_2) = t_2$, by Proposition 4.2.2, there exist $T_2 \in \mathrm{GL}_n(\mathbb{F}_q)$ and $X_2' \in \mathbb{F}_{q^m}^{k\times t_2}$ such that $X_2T_2 = \left(X_2' \mid 0_{k\times(n-t_2)}\right)$. By defining an extended transformation matrix $\bar{T} = \left(\begin{smallmatrix} I_\ell & 0 \\ 0 & T_2 \end{smallmatrix}\right) \in \mathrm{GL}_{n+\ell}(\mathbb{F}_q)$, we can rewrite the public key as

$$\hat{G} = S(X_1 \mid G + X_2)\bar{T}\bar{T}^{-1}T = S(X_1 \mid GT_2 + X_2T_2)Q,$$

where $Q = \bar{T}^{-1}T \in \mathrm{GL}_{n+\ell}(\mathbb{F}_q)$. Let $G' = GT_2$. Right-multiplying a Gabidulin generator matrix by a non-singular matrix over the base field $\mathbb{F}_q$ yields another Gabidulin generator matrix (see Proposition 4.3.3). Thus, G' generates an $[n, k]$ Gabidulin code.

Decompose G' into $(G_1' \mid G_2')$, where G_1' contains the first t_2 columns and G_2' contains the remaining $n - t_2$ columns. Then G_2' acts as a generator matrix for an $[n - t_2, k]$ Gabidulin code. Substituting this decomposition gives us the following expression:

$$G' + X_2T_2 = (G_1' + X_2' \mid G_2').$$

By defining $X^* := (X_1 \mid G_1' + X_2') \in \mathbb{F}_{q^m}^{k\times(\ell+t_2)}$ and setting $G^* := G_2'$, we obtain the desired form $\hat{G} = S(X^* \mid G^*)Q$. Setting $T^* = Q$ completes the construction.

Finally, the new Gabidulin code generated by G^* has length $n - t_2$ and dimension k. Its error-correcting capability is $t^* = \lfloor \frac{n-t_2-k}{2} \rfloor \geq t - \lceil \frac{t_2}{2} \rceil$. Since the public error-correcting capability is bounded by $\hat{t} \leq t - t_2$, it clearly follows that $t^* \geq \hat{t}$. ∎

This formal reduction demonstrates that attempting to mask the entire Gabidulin code with a full-length distortion matrix offers no real cryptographic advantage over adding some distorted columns.

4.3 Overbeck's attack

The GGPT cryptosystem presented in the previous section was broken by Overbeck [79] himself not long after its introduction. The attack uses a distinguisher for Gabidulin codes to obtain an alternative secret key from the public key $(\hat{G}, \hat{t})$. This distinguisher is based on the following operator.

Definition 4.3.1. Let M be an arbitrary $l \times n$ matrix over $\mathbb{F}_{q^m}$ and $f \in \mathbb{N}$. Then, define the matrix operator $\Lambda_f \colon \mathbb{F}_{q^m}^{l\times n} \to \mathbb{F}_{q^m}^{((f+1)\cdot l)\times n}$ by

$$\Lambda_f(M) := \begin{pmatrix} M \\ M^{[1]} \\ \vdots \\ M^{[f]} \end{pmatrix},$$

where $M^{[j]}$ denotes the result of raising every element of M to the power q^j, for every $1 \leq j \leq f$.

Lemma 4.3.2. *For any $A_1, A_2 \in \mathbb{F}_{q^m}^{l\times s}$ and $B \in \mathbb{F}_{q^m}^{s\times n}$, and for any $\alpha, \beta \in \mathbb{F}_q$, it holds that*

$$(\alpha A_1 + \beta A_2)^{[i]} = \alpha A_1^{[i]} + \beta A_2^{[i]} \quad \textit{and} \quad (A_1B)^{[i]} = A_1^{[i]}B^{[i]}.$$

In particular, if $S \in \mathrm{GL}_n(\mathbb{F}_{q^m})$ then also $S^{[i]} \in \mathrm{GL}_n(\mathbb{F}_{q^m})$.

Proof. The map $x \mapsto x^{q^i}$ is an automorphism of the finite field $\mathbb{F}_{q^m}$ for any $i \geq 0$. Therefore, it distributes over both addition and multiplication: $(x+y)^{[i]} = x^{[i]} + y^{[i]}$ and $(xy)^{[i]} = x^{[i]}y^{[i]}$. Furthermore, by the properties of finite fields, any element $\alpha \in \mathbb{F}_q$ satisfies $\alpha^q = \alpha$, which implies $\alpha^{[i]} = \alpha$. By applying these properties component-wise to the matrices, we obtain the first two equalities.

Additionally, if $S \in \mathrm{GL}_n(\mathbb{F}_{q^m})$, there exists an inverse S^{-1} such that

$$SS^{-1} = I_n \quad \Longrightarrow \quad (SS^{-1})^{[i]} = I_n^{[i]}.$$

Because the identity matrix I_n consists solely of elements $0, 1 \in \mathbb{F}_q$, we have $I_n^{[i]} = I_n$. Using the multiplicative property established above, we obtain $S^{[i]}(S^{-1})^{[i]} = I_n$. Thus, $S^{[i]}$ is non-singular and its inverse is exactly $(S^{-1})^{[i]}$. ∎

An interesting consequence of the previous lemma is that right-multiplying the generator matrix of a Gabidulin code by a non-singular matrix over the base field $\mathbb{F}_q$ yields a valid generator matrix for a Gabidulin code with the same parameters.

Proposition 4.3.3. *Let $G \in \mathbb{F}_{q^m}^{k\times n}$ be the generator matrix of an $[n,k]$ Gabidulin code defined by the generator vector $\mathbf{g} \in \mathbb{F}_{q^m}^n$. For any non-singular matrix $T \in \mathrm{GL}_n(\mathbb{F}_q)$, the product GT is a valid generator matrix for the $[n,k]$ Gabidulin code defined by the transformed vector $\mathbf{g}T$.*

Proof. By definition, the j-th row of the generator matrix G is given by the Frobenius power $\mathbf{g}^{[j-1]}$. Using Lemma 4.3.2, and the fact that T has entries strictly in $\mathbb{F}_q$, the j-th row of the product GT evaluates to

$$\mathbf{g}^{[j-1]}T = \mathbf{g}^{[j-1]}T^{[j-1]} = (\mathbf{g}T)^{[j-1]}.$$

This shows that the rows of GT are exactly the successive Frobenius powers of the vector $\mathbf{g}T$. Furthermore, because T is a full-rank matrix over $\mathbb{F}_q$ and the components of $\mathbf{g}$ are linearly independent over $\mathbb{F}_q$, the components of the transformed vector $\mathbf{g}T$ must also remain linearly independent over $\mathbb{F}_q$. Therefore, GT satisfies the definition of a generator matrix for the Gabidulin code generated by $\mathbf{g}T$. ∎

When we apply the operator Λ_f directly to a Gabidulin code, its effect is given by the following lemma. We include a short proof, because the original source lacks one.

Lemma 4.3.4 ([79, Lemma 1]). *If $M \in \mathbb{F}_{q^m}^{l\times n}$ defines an $[n,k]$ Gabidulin code with generator vector $\mathbf{g}$ and $f \leq n-k-1$, then the subspace spanned by the rows of $\Lambda_f(M)$ defines the $[n,k+f]$ Gabidulin code with generator vector $\mathbf{g}$.*

Proof. Let $\mathcal{C}$ be the $[n,k]$ Gabidulin code generated by the vector $\mathbf{g}$. By definition, $\mathcal{C}$ is the subspace spanned by a generator matrix

$$G = \begin{pmatrix} \mathbf{g} \\ \mathbf{g}^{[1]} \\ \vdots \\ \mathbf{g}^{[k-1]} \end{pmatrix}.$$

Since M defines the same code $\mathcal{C}$, its rows must span the same subspace as G. Therefore, there exists a full-rank matrix $A \in \mathbb{F}_{q^m}^{l\times k}$ such that $M = AG$.

Consider the j-th block of $\Lambda_f(M)$, denoted as $M^{[j]}$. Applying the Frobenius automorphism, we obtain, using Lemma 4.3.2,

$$M^{[j]} = (AG)^{[j]} = A^{[j]}G^{[j]}.$$

Since A has full rank, its image under the Frobenius automorphism $A^{[j]}$ also has full rank. Therefore, the row space of $M^{[j]}$ is identical to the row space of $G^{[j]}$.

The operator $\Lambda_f(M)$ stacks the matrices $M^{[0]}, M^{[1]}, \ldots, M^{[f]}$. The subspace spanned by $\Lambda_f(M)$ is therefore the sum of the subspaces spanned by each block:

$$\operatorname{span}(\Lambda_f(M)) = \sum_{j=0}^{f} \operatorname{span}(G^{[j]}) = \operatorname{span}\left(\bigcup_{j=0}^{f} \{\mathbf{g}^{[j]}, \ldots, \mathbf{g}^{[j+k-1]}\} \right).$$

Observing the indices, this union covers the range from 0 to $k + f - 1$. This set of vectors forms the canonical basis for the $[n, k+f]$ Gabidulin code generated by $\mathbf{g}$. Thus, the subspace spanned by $\Lambda_f(M)$ defines this code. ∎

While the previous lemma establishes a definitive structural property for Gabidulin codes, the GGPT public key $\hat{G}$ is masked by random matrices. To analyse the GGPT cryptosystem, two probabilistic assumptions regarding how random subcodes and fully random matrices behave under the Λ_f operator are introduced next.

Assumption 3 ([79, Assumption 1]). Let $M \in \mathbb{F}_{q^m}^{l \times n}$ define a random l-dimensional subcode (with $l > 1$) of an $[n, k]$ Gabidulin code over $\mathbb{F}_{q^m}$ with generator vector $\mathbf{g}$. Then, with probability $\mathcal{P}_1 \geq 1 - q^{-m}$, $\Lambda_f(M)$ defines a $\min\{k + f, (f+1) \cdot l\}$-dimensional subcode of the $[n, k+f]$ Gabidulin code with generator vector $\mathbf{g}$.

Assumption 4 ([79, Assumption 2]). Let $M \in \mathbb{F}_{q^m}^{l \times n}$ be a random matrix of full rank over $\mathbb{F}_{q^m}$ and of full column rank over $\mathbb{F}_q$. Then, $\Lambda_f(M)$ has rank $\min(n, f \cdot l)$ with probability $\mathcal{P}_2 \geq 1 - q^{-(m-1)}$.

The validity of these assumptions were studied by Overbeck [79] with both mathematical reasoning and empirical results. For parameters relevant to a practical cryptosystem (and thus attacks), the probabilities $\mathcal{P}_1$ and $\mathcal{P}_2$ approach 1.

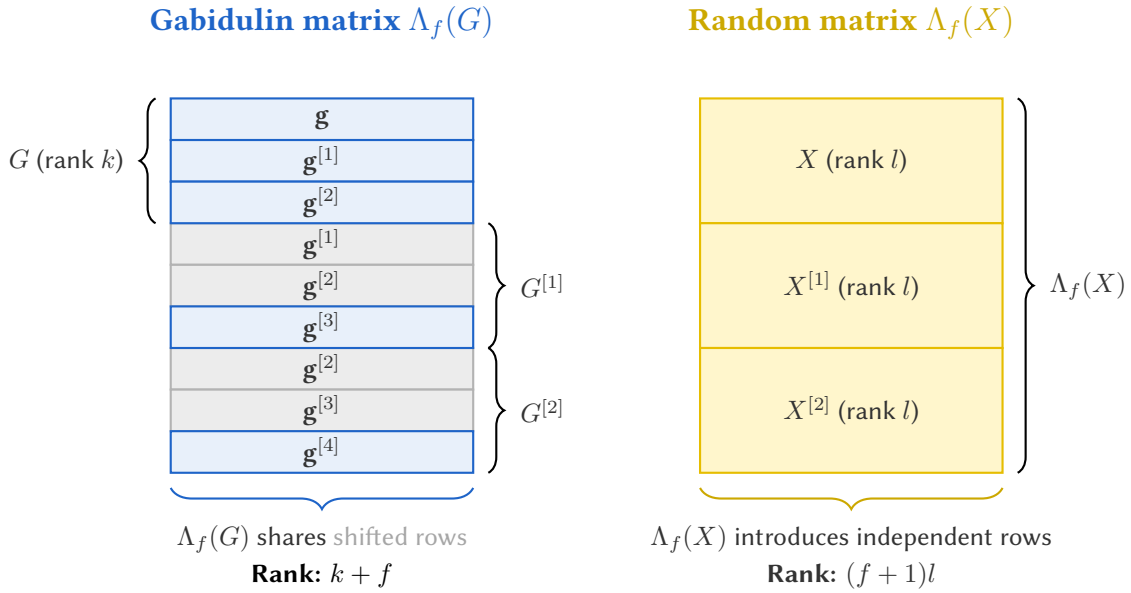


Figure 4.1: Visualisation of the Λ_f operator (illustrated for $k = 3$ and $f = 2$) applied to a Gabidulin (generator) matrix versus a random matrix. The differing rank growth rates expose the underlying algebraic structure.

To recover an alternative secret key from the public key $(\hat{G}, \hat{t})$ of an instance of the GGPT cryptosystem, we observe the behaviour of the matrix $\Lambda_f(\hat{G})$ (see Figure 4.1). The rank of the random masking matrices expands rapidly under the Λ_f operator, whereas the rank of the Gabidulin generator matrix

increases by exactly 1 per application (see Lemma 4.3.4). Therefore, if the difference in rank between $\Lambda_f(\hat{G})$ and $\Lambda_{f+1}(\hat{G})$ is exactly 1 for some integer f, we assume that $\Lambda_f(\hat{G})$ is closely related, in the sense of the following theorem, to a Gabidulin code. The rapid growth of the random distortion has effectively saturated in the ambient space, isolating the underlying algebraic structure.

Theorem 4.3.5 ([79, Theorem 2]). *Let $(\hat{G}, \hat{t})$ be the public key of an instance of the GGPT cryptosystem with parameters q, m, n, k, ℓ, and s. Furthermore, let (G, S, T) be the corresponding secret key. Then, for $0 \leq f \leq n - \ell - k - 1$, there exists a dual matrix of $\Lambda_f(\hat{G})$ of the form*

$$\Lambda_f(\hat{G})^{\perp} = \begin{bmatrix} 0 & H_f^{\top} \\ B_1 & B_2 \end{bmatrix} (T^{-1})^{\top} \in \mathbb{F}_{q^m}^{(n-\ell-k-f+l)\times n}, \tag{4.5}$$

where $H_f \in \mathbb{F}_{q^m}^{(n-\ell)\times(n-\ell-k-f)}$ is the parity-check matrix of an $[n-\ell, k+f]$ Gabidulin code $\mathcal{C}_f$, B_1 is an $l \times \ell$ matrix with $0 \leq l \leq \ell$, and B_2 is some $l \times (n-\ell)$ matrix.

The previous theorem establishes that the parity-check matrix H_f of a Gabidulin code is embedded within the dual space of $\Lambda_f(\hat{G})$. The ideal scenario for an attacker occurs when the number of extra random rows, denoted by l, is exactly zero. When the dual matrix attains this minimal theoretical rank, the extra blocks B_1 and B_2 vanish entirely. This minimal rank case directly allows the polynomial-time extraction of an equivalent private key, forming the core of the attack.

Theorem 4.3.6 ([79, Theorem 3]). *Let $(\hat{G}, \hat{t})$ be the public key of an instance of the GGPT cryptosystem. Given an integer $f \leq n-\ell-k-1$ such that the rank of $\Lambda_f(\hat{G})^{\perp}$ achieves its minimal value of $n-\ell-k-f$, an alternative secret key corresponding to $\hat{G}$ can be recovered in $\mathcal{O}(n^3)$ operations over $\mathbb{F}_{q^m}$.*

Proof. With the conditions in the statement, using Theorem 4.3.5, we know that

$$\Lambda_f(\hat{G})^{\perp} = \left(0 \mid H_f^{\top}\right) (T^{-1})^{\top} \in \mathbb{F}_{q^m}^{(n-\ell-k-f)\times n},$$

where H_f is a parity-check matrix of an $[n-\ell, k+f]$ Gabidulin code $\mathcal{C}_f$. This dual matrix $\Lambda_f(\hat{G})^{\perp}$ can be computed via Gaussian elimination in $\mathcal{O}(n^3)$ operations over $\mathbb{F}_{q^m}$ from $\Lambda_f(\hat{G})$. Note that $H_f^{\top}$ and T remain unknown.

However, because of the specific form of $\Lambda_f(\hat{G})^{\perp}$, we are able to identify a subset N_1 consisting of $n-\ell$ columns of $\Lambda_f(\hat{G})^{\perp}$ that are linearly independent over $\mathbb{F}_q$. Without loss of generality, assume these are the last $n-\ell$ columns, meaning $N_1 = \{\ell+1, \ldots, n\}$. Since these columns form an $\mathbb{F}_q$-basis for the column space of $\Lambda_f(\hat{G})^{\perp}$, they represent a valid check matrix for the column-scrambled instance $\mathcal{C}_f$. We can therefore define an alternative (transposed) check matrix as $\bar{H}_f^{\top} := \Lambda_f(\hat{G})^{\perp}_{N_1}$.

Then $\Lambda_f(\hat{G})^{\perp}$ consists of $\bar{H}_f^{\top}$ multiplied by an invertible scrambler matrix over $\mathbb{F}_q$. Thus, the remaining ℓ columns of $\Lambda_f(\hat{G})^{\perp}$ (the first ℓ columns) must be $\mathbb{F}_q$-linear combinations of the columns in N_1. Therefore, there exists a matrix $\tilde{T} \in \mathbb{F}_q^{\ell\times(n-\ell)}$ such that

$$\Lambda_f(\hat{G})^{\perp}_{\{1,\ldots,\ell\}} = \bar{H}_f^{\top} \cdot \tilde{T}^{\top}.$$

By solving this system of linear equations over $\mathbb{F}_q$, we are able to compute $\tilde{T}$. With $\tilde{T}$ recovered, we construct

$$\bar{T} := \begin{pmatrix} I_\ell & -\tilde{T} \\ 0 & I_{n-\ell} \end{pmatrix} \in \mathbb{F}_q^{n\times n} \quad \Longrightarrow \quad \bar{T}^{-1} = \begin{pmatrix} I_\ell & \tilde{T} \\ 0 & I_{n-\ell} \end{pmatrix}.$$

Right-multiplying the dual matrix $\Lambda_f(\hat{G})^{\perp} = \left(\bar{H}_f^{\top}\tilde{T}^{\top} \mid \bar{H}_f^{\top}\right)$ by $\bar{T}^{\top} = \left(\begin{smallmatrix} I_\ell & 0 \\ -\tilde{T}^{\top} & I_{n-\ell} \end{smallmatrix}\right)$ clears the first ℓ columns:

$$\Lambda_f(\hat{G})^{\perp}\bar{T}^{\top} = \left(\bar{H}_f^{\top}\tilde{T}^{\top} - \bar{H}_f^{\top}\tilde{T}^{\top} \mid \bar{H}_f^{\top}\right) = \left(0_{(n-\ell-k-f)\times\ell} \mid \bar{H}_f^{\top}\right). \tag{4.6}$$

At this point, we know $\bar{H}_f$ and may recover $\bar{H}_0$ from it, because the smaller matrix $\bar{H}_f$ corresponds exactly to the first $n-\ell-k-f$ rows of $\bar{H}_0$. So, by taking the last row of the already computed $\bar{H}_f$ and successively applying the Frobenius automorphism to the last row, we can compute the remaining f missing rows.

Applying the inverse scrambler matrix $\bar{T}^{-1}$ to the public key $\hat{G}$ gives

$$\hat{G}\bar{T}^{-1} = \left(\hat{G}_{\{1,\dots,\ell\}} \mid \hat{G}_{\{1,\dots,\ell\}}\tilde{T} + \hat{G}_{\{\ell+1,\dots,n\}}\right).$$

Define $Y := \hat{G}_{\{1,\dots,\ell\}}\tilde{T} + \hat{G}_{\{\ell+1,\dots,n\}}$. Then, we obtain

$$\begin{aligned}\Lambda_f(\hat{G})(\Lambda_f(\hat{G})^{\perp})^{\top} &= \Lambda_f(\hat{G})\bar{T}^{-1}\begin{pmatrix}0_{\ell\times(n-\ell-k-f)}\\ \bar{H}_f\end{pmatrix} \\ &= \left(\Lambda_f\left(\hat{G}_{\{1,\dots,\ell\}}\right)\tilde{T} + \Lambda_f\left(\hat{G}_{\{\ell+1,\dots,n\}}\right)\right)\bar{H}_f \\ &= \Lambda_f(Y)\bar{H}_f. \qquad \text{(Lemma 4.3.2)}\end{aligned} \tag{4.6}$$

By the definition of the dual space, we know $\Lambda_f(\hat{G})(\Lambda_f(\hat{G})^{\perp})^{\top} = 0$, so $\Lambda_f(Y)\bar{H}_f = 0$.

In particular, this shows that the rows of Y are orthogonal to $\bar{H}_f$ and so, because $\bar{H}_0$ is constructed from $\bar{H}_f$, we have $Y\bar{H}_0 = 0$. We conclude that $\left(0_{(n-\ell-k)\times\ell} \mid \bar{H}_0^{\top}\right)$ is in the dual space of $\hat{G}\bar{T}^{-1}$.

This guarantees that the last $n-\ell$ columns of the matrix $\hat{G}\bar{T}^{-1}$ define an $[n-\ell, k]$ Gabidulin code. Hence, $\bar{T}$ serves as an alternative column scrambler. We can then use the techniques from [27, Section 5.2] to decompose the last $n-\ell$ columns of $\hat{G}\bar{T}^{-1}$ into an alternative row scrambler $\bar{S}$ and an equivalent Gabidulin generator matrix $\bar{G}$ in $\mathcal{O}(k^3)$ operations. Together, the components $(\bar{G}, \bar{S}, \bar{T})$ form a valid equivalent private key that completely breaks the system. ■

In scenarios where the rank of $\Lambda_f(\hat{G})^{\perp}$ is strictly larger than the minimal value $n-\ell-k-f$, the dual matrix contains additional random rows, denoted as B_1 and B_2 in Equation (4.5), which prevent the direct algebraic construction described in the previous proof. To overcome this, the attacker guesses a set N_1 of $n-\ell$ rows such that $\left[0 \mid H_f\right](T^{-1})^{\top}$ restricted to the rows N_1 has full column rank over the base field $\mathbb{F}_q$. Then, the matrix $(\Lambda_f(\hat{G})^{\perp}_{N_1})^{\top}$ corresponds to a so-called Niederreiter instance of the GPT cryptosystem as long as $k+f-l>1$.

The Niederreiter variant uses a public parity-check matrix $\hat{H}$ derived from a subcode of a Gabidulin code. Overbeck demonstrated that this variant is also vulnerable to a polynomial-time attack using the Λ_f operator [79]. By extending the public check matrix to $\Lambda_f(\hat{H})$, the attacker can recover the secret generator vector $\mathbf{g}$ with high probability. Once $\mathbf{g}$ is identified, an equivalent check matrix $\bar{H}$ is constructed, allowing the attacker to solve for the secret scrambler matrix S and recover an alternative private key. By employing this Niederreiter solver as a subroutine, the GGPT cryptosystem can be broken even when the minimal rank condition is not initially met.

To estimate the success probability of the attack, the rank of $\Lambda_f(\hat{G})^{\perp}$ is determined under Assumption 4.

Theorem 4.3.7 ([79, Theorem 4]). *Let $(\hat{G}, \hat{t})$ be the public key of an instance of the GGPT cryptosystem with parameters q, m, n, k, ℓ, and s. If Assumption 4 holds for $s\times\ell$ matrices over $\mathbb{F}_{q^m}$, then the rank of the dual matrix of $\Lambda_f(\hat{G})$ is at most $R = n-k-f-\min\{\ell, fs\}$ with probability $\mathcal{P}_2$.*

For the specific case where $s=1$, Assumption 4 holds perfectly and the conditions for a direct polynomial-time extraction (where the rank of the dual matrix $\Lambda_f(\hat{G})^{\perp}$ achieves its minimal value) are fulfilled with a probability of 1. For instances where the distortion rank is higher ($s>1$), computer experiments show that the rank of $\Lambda_f(\hat{G})$ heavily depends on the construction of the distortion matrix X. Specifically, for randomly generated distortion matrices, the rank of $\Lambda_f(\hat{G})$ almost always expands to either $k+f+\ell$ or $k+f+(s+1)\cdot\min(f,s)+s\cdot\max(0,f-s)$. While certain highly structured choices

for X can artificially force the rank of the dual matrix $\Lambda_f(\hat{G})^\perp$ to reach its theoretical upper bound, thus rendering the system more secure at first glance, doing so inherently sacrifices necessary degrees of freedom in the private key selection, making such countermeasures cryptographically unsatisfying.

The primary strength of this structural attack lies in its high efficiency: when the appropriate rank conditions are met (meaning the rank of $\Lambda_f(\hat{G})^\perp$ drops to its minimal bound of $n-\ell-k-f$), an alternative private key can be reconstructed in $\mathcal{O}(m^5)$ operations over the base field $\mathbb{F}_q$. This polynomial-time complexity completely compromises any parameter set where there exists an integer $f \leq n-\ell-k-1$ satisfying the inequality $\ell \leq fs$. Consequently, all instances of the original GPT cryptosystem are fundamentally broken, as they implicitly fall within these theoretically insecure bounds. Experimental results underscore the practical impact of this vulnerability, executing successfully on standard hardware from two decades ago in mere hours and did not fail for any random instance of the original GPT cryptosystem [79].

To formulate potentially secure instances of the GGPT variant against this specific attack, parameters must be carefully chosen to prevent the rank conditions for $\Lambda_f(\hat{G})^\perp$ from ever being satisfied. The cryptosystem avoids the vulnerability if the inequality $\ell - fs > f + k$ holds strictly for all possible choices of f. This requirement can be satisfied by selecting a distortion rank s bounded by $s \leq \frac{2\ell-n}{n-\ell-k}$.

However, in 2018, Horlemann-Trautmann et al. [42] successfully extended Overbeck's framework to compromise these supposedly secure parameter sets as well. Because it was a matter of time before such work would be presented, Gabidulin [26] proposed an alternative mitigation strategy in 2008: employing a column scrambler defined over the extension field $\mathbb{F}_{q^m}$ rather than the base field $\mathbb{F}_q$. The cryptanalysis of this specific variant forms the primary focus of the following section.

4.4 Extensions of Overbeck's attack

Before considering the GGPT cryptosystem with a column scrambler T over $\mathbb{F}_{q^m}$ rather than $\mathbb{F}_q$, we present a technical lemma regarding the inverse of a specific matrix form that will be useful later.

Lemma 4.4.1 ([77, Lemma 2]). *Let $P = \left(\begin{smallmatrix} A & 0 \\ C & D \end{smallmatrix}\right)$ where A and D are square matrices. Then, P is non-singular if and only if A and D are non-singular. In that case, the inverse of P equals*

$$P^{-1} = \begin{pmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{pmatrix}.$$

Proof. First, assume A and D are non-singular, meaning their inverses A^{-1} and D^{-1} exist. Let Q be the proposed inverse matrix $\left(\begin{smallmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{smallmatrix}\right)$. We verify:

$$PQ = \begin{pmatrix} A & 0 \\ C & D \end{pmatrix}\begin{pmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{pmatrix} = \begin{pmatrix} AA^{-1} & 0 \\ CA^{-1} - DD^{-1}CA^{-1} & DD^{-1} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}.$$

A similar calculation verifies that $QP = I$. Since there exists a matrix Q such that $PQ = QP = I$, the matrix P is non-singular by definition, and its inverse is exactly $P^{-1} = Q$.

Conversely, assume P is non-singular. Because P is a (block) lower triangular matrix, its determinant is given by the product of the determinants of its diagonal blocks:

$$\det(P) = \det(A)\det(D).$$

Since P is invertible, its determinant must be non-zero. This means that $\det(A) \neq 0$ and $\det(D) \neq 0$. Therefore, both A and D must be non-singular matrices. ∎

We are now ready to cryptanalyse the GGPT cryptosystem with a column scrambler defined over the extension field rather than the base field.

4.4.1 Original reparation by Gabidulin

In this section, we consider the "reparation" given in [26], which was the first to consider a column scrambler matrix over the extension field. To establish notation, we give the key generation and decryption steps explicitly. The other steps remain unchanged.

- **Key generation:** First, generate the following matrices over $\mathbb{F}_{q^m}$:
  - G: $k \times n$ generator matrix of an $[n, k]$ Gabidulin code $\mathcal{G}$ with error-correcting capability t;
  - S: $k \times k$ random non-singular matrix (the row scrambler);
  - X: $k \times \ell$ random matrix (the distortion matrix);
  - T: $(n + \ell) \times (n + \ell)$ random non-singular matrix (the column scrambler), such that there exist $Q_{11} \in \mathbb{F}_{q^m}^{\ell \times \ell}$, $Q_{12} \in \mathbb{F}_{q^m}^{\ell \times n}$, $Q_{21} \in \mathbb{F}_{q^m}^{n \times \ell}$, $Q_{22} \in \mathbb{F}_q^{n \times n}$ with $\mathrm{rk}_q(Q_{12}) = s < t$ so that

$$T^{-1} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}. \tag{4.7}$$

  Then, construct the public matrix $\hat{G}$:

$$\hat{G} = S\,(X \mid G)\,T.$$

- **Decryption:** To decode a ciphertext $\mathbf{c} = \mathbf{x}\hat{G} + \mathbf{e}$, where $\mathbf{e} \in \mathbb{F}_{q^m}^{n+\ell}$ is a random error vector of rank weight at most $\hat{t} = t - s$, first remove the column scrambler to obtain $\mathbf{c}' = \mathbf{c}T^{-1} = \mathbf{x}S(X \mid G) + \mathbf{e}T^{-1}$. Suppose that the error term is partitioned as $\mathbf{e} = (\mathbf{e}_1 \mid \mathbf{e}_2)$ where $\mathbf{e}_1 \in \mathbb{F}_{q^m}^{\ell}$ and $\mathbf{e}_2 \in \mathbb{F}_{q^m}^{n}$. The scrambled error term expands to

$$\mathbf{e}T^{-1} = (\mathbf{e}_1 Q_{11} + \mathbf{e}_2 Q_{21} \mid \mathbf{e}_1 Q_{12} + \mathbf{e}_2 Q_{22}).$$

  Applying the triangle inequality to the second part gives

$$\mathrm{rk}_q(\mathbf{e}_1 Q_{12} + \mathbf{e}_2 Q_{22}) \leq \mathrm{rk}_q(\mathbf{e}_1 Q_{12}) + \mathrm{rk}_q(\mathbf{e}_2 Q_{22}).$$

  Because the rank of Q_{12} is exactly s, multiplying it by $\mathbf{e}_1$ cannot yield a rank larger than s (see Corollary 4.2.3). Furthermore, because Q_{22} has entries strictly in the base field $\mathbb{F}_q$, multiplying $\mathbf{e}_2$ by Q_{22} does not increase its rank beyond the total added error $\hat{t} \leq t - s$. Thus,

$$\mathrm{rk}_q(\mathbf{e}_1 Q_{12} + \mathbf{e}_2 Q_{22}) \leq s + (t - s) = t.$$

  Therefore, the plaintext $\mathbf{x}$ is successfully recovered by applying the decoding algorithm to the last n components of $\mathbf{c}T^{-1}$.

The requirement that T^{-1} possesses the specific block structure described above is necessary. If T^{-1} were a totally random non-singular matrix over $\mathbb{F}_{q^m}$, there is a high probability that the rank of $\mathbf{e}T^{-1}$ exceeds the error-correcting capability t of the Gabidulin code generated by G. This would make decryption impossible, also for the legitimate receiver. Thus, forcing the elements of Q_{22} to lie within the base field $\mathbb{F}_q$ and restricting the rank of Q_{12} is the only way to keep the cryptosystem usable. It is also this structure that will be exploited.

In [77], it was proven that this cryptosystem is still vulnerable to Overbeck's attack by considering the dual of the code generated by $\Lambda_{n-s-k-1}(\hat{G})$. Before proving that an alternative distortion, generator and column scrambler matrix can be found that allows for decryption, we require the following lemma.

Lemma 4.4.2 ([77, Lemma 3]). *There exist matrices $T_{11} \in \mathrm{GL}_{\ell+s}(\mathbb{F}_{q^m})$, $T_{21} \in \mathbb{F}_{q^m}^{(n-s)\times(\ell+s)}$, and $T_{22} \in \mathrm{GL}_{n-s}(\mathbb{F}_q)$ such that*

$$T = \begin{pmatrix} I_\ell & 0 \\ 0 & L \end{pmatrix} \begin{pmatrix} T_{11} & 0 \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} I_\ell & 0 \\ 0 & R \end{pmatrix},$$

with L and R belonging to $\mathrm{GL}_n(\mathbb{F}_q)$.

Proof. By assumption, $\mathrm{rk}_q(Q_{12}) = s < t$. So, there exist $R \in \mathrm{GL}_n(\mathbb{F}_q)$ and $Q_{12}^* \in \mathbb{F}_{q^m}^{\ell\times s}$ such that $Q_{12}R = (Q_{12}^* \mid 0)$ (see Proposition 4.2.2). We partition $Q_{22}R = (Q'_{22} \mid Q'_{23})$, where $Q'_{22} \in \mathbb{F}_q^{n\times s}$ and $Q'_{23} \in \mathbb{F}_q^{n\times(n-s)}$. Since Q'_{23} has $n-s$ columns, its rank is at most $n-s$. Therefore, there exists $L \in \mathrm{GL}_n(\mathbb{F}_q)$ such that $LQ'_{23} = \begin{pmatrix} 0 \\ Q_{23}^* \end{pmatrix}$ with $Q_{23}^* \in \mathbb{F}_q^{(n-s)\times(n-s)}$ (analogous to Proposition 4.2.2).

Using Equation (4.7), we can rewrite the transformed inverse matrix as

$$\begin{aligned} \begin{pmatrix} I_\ell & 0 \\ 0 & L \end{pmatrix} T^{-1} \begin{pmatrix} I_\ell & 0 \\ 0 & R \end{pmatrix} &= \begin{pmatrix} I_\ell & 0 \\ 0 & L \end{pmatrix} \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} I_\ell & 0 \\ 0 & R \end{pmatrix} \\ &= \begin{pmatrix} Q_{11} & Q_{12}^* & 0 \\ LQ_{21} & LQ'_{22} & LQ'_{23} \end{pmatrix}. \end{aligned}$$

Then regroup the entries into new blocks:

$$\begin{pmatrix} I_\ell & 0 \\ 0 & L \end{pmatrix} T^{-1} \begin{pmatrix} I_\ell & 0 \\ 0 & R \end{pmatrix} = \begin{pmatrix} Q''_{11} & 0 \\ Q''_{21} & Q_{23}^* \end{pmatrix}, \tag{4.8}$$

where $Q''_{11} \in \mathbb{F}_{q^m}^{(\ell+s)\times(\ell+s)}$ and $Q''_{21} \in \mathbb{F}_{q^m}^{(n-s)\times(\ell+s)}$. Because this matrix is formed by the product of invertible matrices, the entire left-hand side is invertible. Therefore, by Lemma 4.4.1, the blocks Q''_{11} and Q_{23}^* must also be invertible.

Finally, inverting both sides of Equation (4.8) preserves this structure, again by Lemma 4.4.1. Relabelling the resulting blocks yields the target decomposition for T, completing the proof. ∎

Remark 4.4.3. The proof of Lemma 4.4.2 explicitly assumes that $\mathrm{rk}_q(Q_{12}) = s$. However, the structural decomposition holds true even under the relaxed assumption that $\mathrm{rk}_q(Q_{12}) \leq s$. If the actual rank is $r < s$, Proposition 4.2.2 still guarantees the existence of a transformation that "compresses" the column space. We can construct the $\ell \times s$ block Q_{12}^* by taking the r linearly independent columns and adding $s-r$ linearly dependent (or all-zero) columns. This still guarantees at least $n-s$ zero columns on the right, fully preserving the required matrix (block) structure during the regrouping step. Additionally, because Q_{12} is an $\ell \times n$ matrix by definition, its rank is fundamentally bounded by its row dimension. Therefore, the parameter s is necessarily less than or equal to ℓ.

The cryptosystem with the column scrambler defined in this section is formally broken by the following theorem. It essentially shows the existence of an alternate form of the public key $\hat{G}$ with a column scrambler in the base field $\mathbb{F}_q$ and for which the associated Gabidulin code can still correct sufficient errors to allow for decryption.

Theorem 4.4.4 ([77, Proposition 8]). *There exist matrices $X^* \in \mathbb{F}_{q^m}^{k\times(\ell+s)}$, $T^* \in \mathrm{GL}_{n+\ell}(\mathbb{F}_q)$, and a generator matrix G^* that defines an $[n-s, k]$ Gabidulin code such that*

$$\hat{G} = S\,(X^* \mid G^*)\,T^*.$$

Furthermore, the error-correcting capability t^* of the Gabidulin code generated by G^* is greater than or equal to the public error-correcting capability $\hat{t} = t - s$.

Proof. We keep the notation established in Lemma 4.4.2. First, we apply the transformation $L \in \mathrm{GL}_n(\mathbb{F}_q)$ to the original Gabidulin generator matrix G, yielding another valid generator matrix GL for an $[n, k]$ Gabidulin code (see Proposition 4.3.3). We partition this transformed matrix as $GL = (G'_1 \mid G'_2)$, where $G'_1 \in \mathbb{F}_{q^m}^{k \times s}$ consists of the first s columns, and $G'_2 \in \mathbb{F}_{q^m}^{k \times (n-s)}$ contains the remaining $n - s$ columns. Crucially, G'_2 itself acts as a generator matrix for an $[n - s, k]$ Gabidulin code.

Now, substitute the decomposed form of T from Lemma 4.4.2 into the public key equation $\hat{G} = S(X \mid G)T$:

$$\hat{G} = S\,(X \mid G) \begin{pmatrix} I_\ell & 0 \\ 0 & L \end{pmatrix} \begin{pmatrix} T_{11} & 0 \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} I_\ell & 0 \\ 0 & R \end{pmatrix}.$$

Next, observe that

$$(X \mid G) \begin{pmatrix} I_\ell & 0 \\ 0 & L \end{pmatrix} = (X \mid GL) = \left(X \mid G'_1 \mid G'_2\right).$$

To simplify the representation, we define $Y := (X \mid G'_1) \in \mathbb{F}_{q^m}^{k \times (\ell+s)}$. Multiplying $(Y \mid G'_2)$ by the middle matrix from the decomposition of T, we obtain

$$\left(Y \mid G'_2\right) \begin{pmatrix} T_{11} & 0 \\ T_{21} & T_{22} \end{pmatrix} = \left(YT_{11} + G'_2T_{21} \mid G'_2T_{22}\right).$$

Define the new distortion matrix $X^* := YT_{11} + G'_2T_{21} \in \mathbb{F}_{q^m}^{k \times (\ell+s)}$ and the new generator matrix $G^* := G'_2T_{22}$. Because $T_{22} \in \mathrm{GL}_{n-s}(\mathbb{F}_q)$ has entries in the base field $\mathbb{F}_q$, G^* is a valid generator matrix for an $[n - s, k]$ Gabidulin code. Finally, by defining the new column scrambler $T^* := \left(\begin{smallmatrix} I_\ell & 0 \\ 0 & R \end{smallmatrix}\right) \in \mathrm{GL}_{n+\ell}(\mathbb{F}_q)$, we can rewrite the public key as

$$\hat{G} = S\,(X^* \mid G^*)\,T^*.$$

To conclude the proof, we evaluate the error-correcting capability t^* of the newly formed Gabidulin code generated by G^*. The original $[n, k]$ Gabidulin code corrects $t = \lfloor \frac{n-k}{2} \rfloor$ errors. The new code has length $n - s$ and dimension k, so its error-correcting capability is

$$t^* = \left\lfloor \frac{n - s - k}{2} \right\rfloor \geq t - \left\lceil \frac{s}{2} \right\rceil.$$

Since the rank s of the noise is positive, we have $t^* \geq t - \lceil \frac{s}{2} \rceil \geq t - s$. Thus, this degraded Gabidulin code is still capable of correcting the maximum possible public error $\hat{t} = t - s$. ∎

The previous theorem guarantees a structural equivalence of the public key that compromises the cryptosystem. In essence, the public key $\hat{G}$ can always be interpreted as a standard GGPT instance where the column scrambler T^* is defined strictly over the base field $\mathbb{F}_q$. This mathematical reduction completely bypasses the column scrambler over an extension field. An attacker can treat $\hat{G}$ as a vulnerable public key and feed it directly into Overbeck's existing structural attack.

Because the equivalent hidden Gabidulin code G^* has dimension k and a reduced length of $n - s$, the attacker applies the operator Λ_f with $f = (n - s) - k - 1$. According to Assumption 4 and Theorem 4.3.5, this specific choice of f causes the dual space $\Lambda_{n-s-k-1}(\hat{G})^\perp$ to collapse to its minimal theoretical rank (with high probability), isolating the parity-check matrix of the code. The attacker then uses Theorem 4.3.6 to recover an equivalent generator matrix for the Gabidulin code generated by G^*. Finally, because the degraded error-correcting capability t^* is strictly larger than the maximum injected error $\hat{t} = t - s$, the attacker can successfully apply a standard Gabidulin decoding algorithm to this recovered code to decrypt any intercepted ciphertext.

4.4.2 Variant by Gabidulin, Rashwan and Honary

In [30], Gabidulin, Rashwan and Honary proposed another variant of the GGPT cryptosystem with the entries of the column scrambler T in the extension field $\mathbb{F}_{q^m}$. Otmani et al. [77] proved that this scheme is actually a special case of the one presented in the previous section, and because of that, it suffers from the same weakness. Therefore, it is also susceptible to Overbeck's attack.

The key generation and decryption steps of the variant by Gabidulin, Rashwan and Honary are as follows.

- **Key generation:** First, generate the following matrices over $\mathbb{F}_{q^m}$:
  - G: $k \times n$ generator matrix of an $[n, k]$ Gabidulin code $\mathcal{G}$ with error-correcting capability t, choose $\hat{t} < t$, and define $a = t - \hat{t}$;
  - S: $k \times k$ random non-singular matrix (the row scrambler);
  - T: $n \times n$ random non-singular matrix (the column scrambler) defined over the extension field $\mathbb{F}_{q^m}$, such that there exist matrices $Q_1 \in \mathbb{F}_{q^m}^{n \times a}$ and $Q_2 \in \mathbb{F}_q^{n \times (n-a)}$ so that

$$T^{-1} = (Q_1 \mid Q_2)\,.$$

  Then, construct the public matrix $\hat{G}$:

$$\hat{G} = SGT.$$

- **Decryption:** To decode a ciphertext $\mathbf{c} = \mathbf{x}\hat{G} + \mathbf{e}$, where $\mathbf{e} \in \mathbb{F}_{q^m}^n$ is a random error vector of rank weight at most $\hat{t}$, first remove the column scrambler to obtain $\mathbf{c}' = \mathbf{c}T^{-1} = \mathbf{x}SG + \mathbf{e}T^{-1}$. The scrambled error term expands to

$$\mathbf{e}T^{-1} = (\mathbf{e}Q_1 \mid \mathbf{e}Q_2)\,.$$

  As before, using Corollary 4.2.3 and the fact that Q_2 has entries in the base field, applying the triangle inequality gives

$$\mathrm{rk}_q(\mathbf{e}T^{-1}) \leq \mathrm{rk}_q(\mathbf{e}Q_1) + \mathrm{rk}_q(\mathbf{e}Q_2) \leq a + \hat{t} = (t - \hat{t}) + \hat{t} = t.$$

  Therefore, the plaintext $\mathbf{x}$ is successfully recovered by applying the decoding algorithm to $\mathbf{c}T^{-1}$.

As established in [77], this formulation essentially functions as a specific instance of Gabidulin's general reparation where no explicit distortion matrix X is added (meaning $\ell = 0$), and the upper-right block Q_{12} of the inverse column scrambler has all its entries in the base field $\mathbb{F}_q$. This allows us to formulate the following corollary, proving that the scheme remains fundamentally broken.

Corollary 4.4.5 ([77, Corollary 2]). *There exist matrices $T^* \in \mathrm{GL}_n(\mathbb{F}_q)$ and $X^* \in \mathbb{F}_{q^m}^{k \times a}$ such that the public key can be rewritten as*

$$\hat{G} = S\,(X^* \mid G^*)\,T^*,$$

where G^* is a generator matrix of an $[n - a, k]$ Gabidulin code. Furthermore, the error-correcting capability t^* of this new Gabidulin code is greater than or equal to the public error-correcting capability $\hat{t}$.

Proof. Apply Theorem 4.4.4 with $\ell = 0$ and $s = a$. Therefore, its error-correcting capability t^* is

$$t^* = \left\lfloor \frac{n - a - k}{2} \right\rfloor \geq t - \left\lceil \frac{a}{2} \right\rceil = \hat{t} + \left\lfloor \frac{a}{2} \right\rfloor.$$

Since $\hat{t} < t$, we have $a \geq 1$, which implies $\lfloor \frac{a}{2} \rfloor \geq 0$. This guarantees that $t^* \geq \hat{t}$. ∎

Because the error-correcting capability of this alternative Gabidulin code is larger than or equal to $\hat{t}$, the maximum rank of an injected error $\mathbf{e}$, an attacker can treat the public key precisely as the standard (and vulnerable) GGPT variant. By applying Λ_f with $f = (n-a)-k-1$, the attacker can extract alternative private key components in polynomial time (Overbeck's attack) to successfully decrypt any intercepted ciphertext.

For the sake of completeness, it is worth mentioning that another generalisation was proposed to improve the scheme's security. In 2009, Gabidulin and his co-authors [30] suggested to use a more complex column scrambler, constructed as the product of two matrices. The intuition was that this extra mixing would further obscure the algebraic structure and disrupt the Frobenius operator. However, as demonstrated in [77, Section 7], this additional protection is useless. The public key structure can again be reduced to a simpler version, remaining vulnerable to Overbeck's attack; the underlying Gabidulin code – though slightly further degraded in its length and parameters – retains enough error-correction capability to correct the injected errors and allow for decryption.

4.4.3 Other variants

An alternative branch of reparations of the GPT cryptosystem focused on the distortion matrix X. The main vulnerability exploited by Overbeck's attack is that applying the operator Λ_{n-k-1} to a public key of length $n+\ell$ yields an expanded code of dimension $n+\ell-1$. Therefore, its dual space has dimension $(n+\ell)-(n+\ell-1)=1$, isolating the underlying parity-check matrix (see Theorem 4.3.6). To counteract this, several works [58, 88] proposed employing special distortion matrices instead of purely random ones. The objective of these specific distortions was to artificially inflate the dimension of $\Lambda_{n-k-1}(\mathcal{C})^{\perp}$ such that it equals some value e, sufficiently large to prevent an attacker from finding an equivalent Gabidulin code through exhaustive search.

Unfortunately, this approach is also broken. Cryptanalytic research, specifically the structural attacks introduced by Horlemann-Trautmann et al. [42], demonstrated that these specialised distortion matrices introduce their own vulnerabilities. One might intuitively suggest combining both defences – using a special distortion matrix alongside a column scrambler over the extension field $\mathbb{F}_{q^m}$ – to build a secure cryptosystem. Yet, as Otmani et al. [77] showed, this combined scheme remains insecure; the column scrambler can be reduced to the base field $\mathbb{F}_q$, leaving the underlying special distortion entirely vulnerable to the extended structural attacks.

The last alternative direction not discussed before is the use of reducible rank codes, first proposed in 2003 [31]. A reducible rank code is constructed by combining multiple rank-metric codes, effectively embedding the structured generator matrices within a larger, less structured matrix. Again, in the context of the GPT cryptosystem, the goal was to hide the algebraic structure of a Gabidulin code, but ultimately the cryptosystem using this concept for its public generator matrix was broken too [81].

Definition 4.4.6 (Reducible rank code). Let $G_1, G_2, \ldots, G_r$ be the generator matrices of $[n_i, k_i]$ rank-metric codes over $\mathbb{F}_{q^m}$ for $i = 1, \ldots, r$. A code $\mathcal{C}$ is called a *reducible rank code* if its generator matrix G can be represented as

$$G = \begin{pmatrix} G_1 & 0 & 0 & \ldots & 0 \\ G_{21} & G_2 & 0 & \ldots & 0 \\ G_{31} & G_{32} & G_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ G_{r1} & G_{r2} & G_{r3} & \ldots & G_r \end{pmatrix}, \tag{4.9}$$

with $G_{ij} \in \mathbb{F}_{q^m}^{k_i \times n_j}$, for $i = 2, \ldots, r$ and $j = 1, \ldots, i-1$. The overall length of the code $\mathcal{C}$ is $n = \sum_{i=1}^{r} n_i$ and its dimension is $k = \sum_{i=1}^{r} k_i$.

Note, however, that this construction solves a fundamental limitation of standard Gabidulin codes. By definition, a standard Gabidulin code requires the components of its generator vector $\mathbf{g}$ to be linearly

independent over the base field $\mathbb{F}_q$. This means that n cannot exceed the extension degree m. If $n > m$, forming such a basis is impossible because the generating vector will contain coordinates that are linearly dependent over $\mathbb{F}_q$. By concatenating multiple smaller Gabidulin codes (where each $n_i \leq m$) on the diagonal blocks, reducible rank codes allow for the construction of optimal maximum rank distance (MRD) codes for lengths $n > m$.

4.5 Consensus on the security of the GPT cryptosystem

The cryptanalysis of the GPT cryptosystem over the past three decades paints a clear picture of the fundamental difficulties in masking highly structured rank-metric codes. As detailed, attempts to prevent structural attacks by employing column scramblers defined over the extension field $\mathbb{F}_{q^m}$ ultimately fail, as they can be mapped back to vulnerable instances defined over the base field $\mathbb{F}_q$. Similarly, attempts to secure the system by employing special distortion matrices introduce other vulnerabilities.

Therefore, the following consensus has emerged within the cryptographic community: all variants of the GPT cryptosystem based on Gabidulin codes are insecure. The root cause of this systemic failure is not the rank metric itself, but rather the intrinsic properties of Gabidulin codes. Their behaviour under the Frobenius operator provides a distinguisher that cannot be effectively hidden by masking techniques. Twisted Gabidulin codes (or AGTG codes) are no option either. In 2023, Couvreur and Zappatore [16] presented an extension of Overbeck's attack that successfully cryptanalyses GPT schemes instantiated with twisted Gabidulin codes (or AGTG codes).

Despite the downfall of the Gabidulin-based GPT scheme, rank-metric cryptography remains a promising candidate for post-quantum standardisation. The hardness of the generic rank syndrome decoding problem remains undisputed. To build secure cryptographic primitives, the focus has shifted towards using code families lacking the algebraic structure that can easily distinguish themselves from random codes.

An example of such codes are low-rank parity-check (LRPC) codes [33]. Introduced by Gaborit, Murat, Ruatta, and Zémor in 2013, LRPC codes act as the rank-metric analogues of classical low-density parity-check (LDPC) codes. LRPC codes are defined by a parity-check matrix whose entries span a low-dimensional subspace. This specific construction gives them a very weak algebraic structure. Because of this, LRPC codes exhibit a behaviour under the Frobenius operator that is indistinguishable from random linear codes, effectively preventing Overbeck's structural attack. However, other challenges arise for this type of code.

Part II

Symmetric cryptography

5 Walsh spectrum of cryptographic functions

The design of a symmetric cryptosystem is complex and combines several different building blocks. Because it is often impossible to analyse the whole system at once, it is important to study the security of these building blocks individually. Vectorial (Boolean) functions play a central role in many block ciphers, i.e. algorithms that encrypt and decrypt data in fixed-size blocks. They are often used as substitution boxes (S-boxes), which are components that substitute input values with different output values to confuse the relationship between the ciphertext and the secret key.

A robust design requires these functions to be resistant to the two most fundamental attacks: differential and linear cryptanalysis. These properties are quantified through differential uniformity and the Walsh spectrum, respectively. While differential uniformity measures a function's vulnerability to differential propagation, the Walsh spectrum characterises its proximity to affine functions, thereby determining its resilience against linear approximations.

In recent work, Bénéteau et al. [7] studied the Walsh spectrum of a special group of functions, called quadratic APN functions, in the binary case. They established a connection between these functions and both vector space partitions and blocking sets in projective space. This chapter extends that framework to fields of odd characteristic, offering new results for the class of crooked quadratic functions. Furthermore, by characterising the algebraic structure underlying the Walsh spectrum as a projective hypersurface regardless of characteristic, we provide a unified geometric perspective that sharpens existing bounds and resolves several open problems originally posed in [7].

Notation. Because we identify the n-dimensional vector space $\mathbb{F}_p^n$ with the extension field $\mathbb{F}_{p^n}$ throughout this chapter, we do not use boldface typography for vectors.

5.1 Introduction

Let n, m, p be positive integers, where p is a prime. An (n, m, p)-function is any function F from $\mathbb{F}_p^n$ to $\mathbb{F}_p^m$. When the parameters n, m, and p are clear from the context, we simply refer to F as a *vectorial function*. In the specific case where $p = 2$, these are known as *vectorial Boolean functions*.

Vectorial functions serve as the fundamental building blocks for many symmetric ciphers. Consequently, the security of the overall cryptographic design relies heavily on the resilience of these components.

5.1.1 Differential cryptanalysis

Differential cryptanalysis studies how specific differences in the input propagate to differences in the output. The resistance of a vectorial function against differential attacks is quantified by the concept of differential uniformity.

Definition 5.1.1 (Difference function). Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ be a vectorial function. The *difference function* $D_{F,a}\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ is defined for any $a \in \mathbb{F}_p^n \setminus \{0\}$ as

$$D_{F,a}(x) = F(x + a) - F(x).$$

Definition 5.1.2 (Differential uniformity). A vectorial function $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ has *differential uniformity* d, if

$$d = \max_{a \in \mathbb{F}_p^n \setminus \{0\}, b \in \mathbb{F}_p^m} |\{ x \in \mathbb{F}_p^n : D_{F,a}(x) = b \}|.$$

A function with differential uniformity 1 is called *perfect nonlinear* (PN) or *planar*, while one with $d = 2$ is called *almost perfect nonlinear* (APN).

A smaller differential uniformity indicates stronger security. Differential cryptanalysis exploits input differences a that yield specific output differences b with high probability; minimising d ensures this probability remains low, thereby preventing a so-called differential attack.

5.1.2 Linear cryptanalysis

In linear cryptanalysis, an attacker approximates the cipher's non-linear behaviour using affine equations. This involves selecting specific subsets of input and output bits and comparing their weighted sums to find hidden correlations. The analysis is geometrically equivalent to projecting the data onto specific directions, with a vulnerable function aligning predictably along one direction. Thus, we must analyse the projections of the function along linear functionals.

To formalise this over the finite field $\mathbb{F}_{p^n}$, we view the field as an n-dimensional vector space over $\mathbb{F}_p$, denoted $\mathbb{F}_p^n$. While abstract vector spaces lack a canonical inner product, the multiplicative structure of the field allows us to define one via the absolute trace function $\mathrm{Tr}_{\mathbb{F}_p} \colon \mathbb{F}_{p^n} \to \mathbb{F}_p$ (see Definition 3.4.10). We define the canonical inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{F}_{p^n}$ as

$$\langle a, x \rangle := \mathrm{Tr}_{\mathbb{F}_p}(ax).$$

Since this bilinear form is non-degenerate, it provides a canonical isomorphism between the field $\mathbb{F}_{p^n}$ (seen as a vector space) and its dual space. This ensures that every linear functional (projection) from $\mathbb{F}_{p^n}$ to $\mathbb{F}_p$ can be uniquely represented as a multiplication by a "direction" element $a \in \mathbb{F}_{p^n}$ followed by the trace function.

This framework is unified: for $p = 2$, it recovers the standard dot product used in [7] (provided the vector space is identified with the field via a self-dual basis, see [55, Definition 2.30]), while for odd p, it remains compatible with the polynomial structure of quadratic functions which we will introduce later.

We use this structure to reduce a vectorial function $F \colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ to its scalar projections. For any output selection mask $b \in \mathbb{F}_p^m$, the *component function* F_b is defined as the projection of F onto b: $F_b(x) = \langle b, F(x) \rangle = \mathrm{Tr}_{\mathbb{F}_p}(bF(x))$. We call F_0 the *trivial component* function.

The magnitude of the Walsh transform measures the correlation between a component function F_b and the linear mask defined by a.

Definition 5.1.3 (Walsh transform and linearity). Let $F \colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ be a vectorial function. Its *Walsh transform* $W_F(b, a)$, for $a \in \mathbb{F}_p^n$ and $b \in \mathbb{F}_p^m$, is defined by

$$W_F(b, a) = \sum_{x \in \mathbb{F}_p^n} \xi_p^{F_b(x) - \langle a, x \rangle},$$

where $\xi_p = e^{2\pi i/p}$. The multiset collecting the values $|W_F(b, a)|$, for all $a \in \mathbb{F}_p^n$ and $b \in \mathbb{F}_p^m$, is called the *extended Walsh spectrum* of F. The *linearity* of F is defined as

$$\mathcal{L}(F) = \max_{a \in \mathbb{F}_p^n, b \in \mathbb{F}_p^m \setminus \{0\}} |W_F(b, a)|.$$

A high magnitude in the spectrum indicates that the function closely resembles an affine function, revealing a strong linear bias that an attacker can exploit.

5.1.3 Quadratic functions

Vectorial functions can be represented as polynomials over finite fields. By restricting the algebraic degree of these polynomials, we obtain specific classes of functions. In this section, we focus on quadratic functions and establish some fundamental results over any field $\mathbb{F}_p$ with p prime.

Definition 5.1.4 (Quadratic function). Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ be a vectorial function. By fixing a basis, F can be represented by an m-tuple of coordinate functions $f_1, \ldots, f_m$, where each $f_i\colon \mathbb{F}_p^n \to \mathbb{F}_p$ is a multivariate polynomial in the ring $\mathbb{F}_p[x_1, \ldots, x_n]$ modulo the ideal $\langle x_1^p - x_1, \ldots, x_n^p - x_n\rangle$.

The *algebraic degree* $\deg(F)$ is defined as the maximum of the total degrees of its coordinate functions f_i. The function F is said to be *affine* when $\deg(F) \leq 1$ and *quadratic* when $\deg(F) = 2$.

The following proposition offers us a useful characterisation. It implies that all component functions F_b, $b \in \mathbb{F}_p^m$, of a quadratic function $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ are the sum of a quadratic form and an affine function over $\mathbb{F}_p$.

Proposition 5.1.5. *The component functions F_b, $b \in \mathbb{F}_p^m$, of a quadratic function $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ can be written as the sum of a quadratic form Q_b and an affine function on $\mathbb{F}_p^n$.*

Proof. Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ be a quadratic function. We can decompose F as $F(x) = F_Q(x) + F_A(x)$, where F_Q contains the strictly quadratic terms and F_A contains the terms of degree at most 1. Likewise, for any $b \in \mathbb{F}_p^m$, we can split the component function $F_b(x) = \langle b, F(x)\rangle$ into $Q_b(x) + L_b(x)$, where $Q_b(x) := \langle b, F_Q(x)\rangle$ and $L_b(x) := \langle b, F_A(x)\rangle$.

To express F_Q explicitly for an input vector $x = (x_1, \ldots, x_n) \in \mathbb{F}_p^n$, consider the purely quadratic part of each coordinate function f_k (see Definition 5.1.4). It consists of a linear combination of degree-two monomials $x_i x_j$ for $1 \leq i \leq j \leq n$, with some scalar coefficients $\gamma_{ij}^{(k)} \in \mathbb{F}_p$. By grouping these scalar coefficients across all m coordinates into vectors $c_{ij} = (\gamma_{ij}^{(1)}, \ldots, \gamma_{ij}^{(m)}) \in \mathbb{F}_p^m$, we can write

$$F_Q(x) = \sum_{1 \leq i \leq j \leq n} c_{ij} x_i x_j.$$

Since F_A is an affine function over $\mathbb{F}_p$ and the inner product is $\mathbb{F}_p$-linear, it follows that $L_b\colon \mathbb{F}_p^n \to \mathbb{F}_p$ is an affine function. To show that Q_b is a quadratic form, note that $Q_b(x) = \sum_{1 \leq i \leq j \leq n} \langle b, c_{ij}\rangle x_i x_j$. Let $a_{ij} = \langle b, c_{ij}\rangle \in \mathbb{F}_p$. Applying a scalar $\lambda \in \mathbb{F}_p$ to any vector $x \in \mathbb{F}_p^n$ scales it component-wise to $\lambda x = (\lambda x_1, \ldots, \lambda x_n)$. Substituting this into our expression, we obtain

$$Q_b(\lambda x) = \sum_{1 \leq i \leq j \leq n} a_{ij} (\lambda x_i)(\lambda x_j) = \lambda^2 \sum_{1 \leq i \leq j \leq n} a_{ij} x_i x_j = \lambda^2 Q_b(x).$$

It remains to verify that the associated form $B_b(x, y) = Q_b(x + y) - Q_b(x) - Q_b(y)$ is bilinear. Expanding this difference, we observe that

$$\begin{aligned} B_b(x, y) &= \sum_{1 \leq i \leq j \leq n} a_{ij} \left[(x_i + y_i)(x_j + y_j) - x_i x_j - y_i y_j\right] \\ &= \sum_{1 \leq i \leq j \leq n} a_{ij} \left[x_i y_j + y_i x_j\right]. \end{aligned}$$

For any scalars $\alpha, \beta \in \mathbb{F}_p$ and vectors $u, v \in \mathbb{F}_p^n$, substituting $x = \alpha u + \beta v$ into the first argument yields

$$\begin{aligned} B_b(\alpha u + \beta v, y) &= \sum_{1 \leq i \leq j \leq n} a_{ij} \left[(\alpha u_i + \beta v_i) y_j + y_i (\alpha u_j + \beta v_j)\right] \\ &= \alpha \sum_{1 \leq i \leq j \leq n} a_{ij} \left(u_i y_j + y_i u_j\right) + \beta \sum_{1 \leq i \leq j \leq n} a_{ij} \left(v_i y_j + y_i v_j\right) \\ &= \alpha B_b(u, y) + \beta B_b(v, y). \end{aligned}$$

Linearity in the second argument follows by symmetry. Therefore, B_b is a bilinear form, confirming that Q_b is indeed a quadratic form for all $b \in \mathbb{F}_p^m$. ∎

Remark 5.1.6. Quadratic functions are often defined immediately as functions whose component functions are the sum of a quadratic form and an affine function. This is indeed an equivalent definition: if every component function has algebraic degree at most two, then specifically the component functions corresponding to the basis vectors of $\mathbb{F}_p^m$ have degree at most two. Since these are precisely the coordinate functions of F, the overall algebraic degree of F is at most two.

The bilinear form associated with the quadratic form can be represented as a matrix. The following lemma proves a fundamental property of these matrices that holds for any characteristic.

Definition 5.1.7 (Homogeneous polynomial). A polynomial $P \in \mathbb{F}_p[x_1, \ldots, x_n]$ is called homogeneous of degree d if every non-zero term has total degree d. Equivalently, for any scalar $\lambda \in \mathbb{F}_p$, it holds that $P(\lambda x_1, \ldots, \lambda x_n) = \lambda^d P(x_1, \ldots, x_n)$.

Lemma 5.1.8. *Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ be a quadratic function. For any $b \in \mathbb{F}_p^m$, let M_b be the $n \times n$ matrix representing the symmetric bilinear form $B_b(x, y)$ associated with the component function F_b. Then the entries of M_b are homogeneous linear polynomials in the coordinates $b_1, \ldots, b_m$ of b.*

Proof. By Proposition 5.1.5, the quadratic form associated with F_b is $Q_b(x) = \langle b, F_Q(x)\rangle$. Its corresponding bilinear form is defined as $B_b(x, y) = Q_b(x + y) - Q_b(x) - Q_b(y)$. Using the $\mathbb{F}_p$-linearity of the inner product, we obtain

$$B_b(x, y) = \langle b, F_Q(x + y) - F_Q(x) - F_Q(y)\rangle.$$

Fix a basis $\{e_1, \ldots, e_n\}$ for $\mathbb{F}_p^n$ and $\{e'_1, \ldots, e'_m\}$ for $\mathbb{F}_p^m$. For any pair of basis vectors e_i and e_j, the vector $K_{ij} = F_Q(e_i + e_j) - F_Q(e_i) - F_Q(e_j)$ is independent of b. Expanding b in the chosen basis as $b = \sum_{k=1}^m b_k e'_k$, and using the $\mathbb{F}_p$-linearity of the inner product, the matrix entry equals

$$(M_b)_{ij} = B_b(e_i, e_j) = \langle b, K_{ij}\rangle = \sum_{k=1}^{m} b_k \langle e'_k, K_{ij}\rangle.$$

Since the inner products $\langle e'_k, K_{ij}\rangle$ are scalars in $\mathbb{F}_p$, each entry $(M_b)_{ij}$ is a homogeneous linear polynomial in the variables $b_1, \ldots, b_m$. ∎

If we look at the difference function (see Definition 5.1.1) of quadratic functions, then it contains only linear and constant terms.

Lemma 5.1.9. *For a quadratic function $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ and any $a \in \mathbb{F}_p^n \setminus \{0\}$, the difference function $D_{F,a}$ can be written as*

$$D_{F,a}(x) = L_a(x) + c_a,$$

where $L_a\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ is a linear map over $\mathbb{F}_p$ and $c_a \in \mathbb{F}_p^m$ is a constant.

Proof. We can decompose F as $F(x) = F_Q(x) + F_A(x)$, where F_Q contains the (strictly) quadratic terms and F_A is an affine function. For any non-zero $a \in \mathbb{F}_p^n$, the difference function of the affine part $D_{F_A,a}(x) = F_A(x + a) - F_A(x)$ evaluates to a constant vector $c_{A,a} \in \mathbb{F}_p^m$. Thus, we only need to evaluate the quadratic part F_Q because $D_{F,a}(x) = D_{F_Q,a}(x) + D_{F_A,a}(x)$.

Following the representation in the proof of Proposition 5.1.5, we can explicitly write the quadratic part as $F_Q(x) = \sum_{1\le i\le j\le n} c_{ij} x_i x_j$, where $c_{ij} \in \mathbb{F}_p^m$ and $x_1, \ldots, x_n$ are the coordinates of $x \in \mathbb{F}_p^n$. For a fixed non-zero $a = (a_1, \ldots, a_n) \in \mathbb{F}_p^n$, we calculate

$$\begin{aligned} D_{F_Q,a}(x) &= \sum_{1\le i\le j\le n} c_{ij}\left[(x_i + a_i)(x_j + a_j) - x_i x_j\right] \\ &= \sum_{1\le i\le j\le n} c_{ij}(a_i x_j + a_j x_i + a_i a_j). \end{aligned}$$

We separate this into a part linearly dependent on the coordinates of x and a constant part:

$$L_a(x) := \sum_{1 \le i \le j \le n} c_{ij}(a_i x_j + a_j x_i) \quad \text{and} \quad c_{Q,a} := \sum_{1 \le i \le j \le n} c_{ij} a_i a_j = F_Q(a).$$

Because a is fixed, $L_a(x)$ is a linear combination of the variables x_k, $k \in \{1, \ldots, n\}$, over $\mathbb{F}_p$, meaning L_a is a linear map from $\mathbb{F}_p^n$ to $\mathbb{F}_p^m$. Setting the overall constant to $c_a = c_{Q,a} + c_{A,a}$, we obtain $D_{F,a}(x) = L_a(x) + c_a$, concluding the proof. ∎

The image set of any difference function $D_{F,a}$ is referred to as a *differential set* of the vectorial function $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ with respect to $a \in \mathbb{F}_p^n \setminus \{0\}$. The so-called crooked property, first introduced in [6] for $p = 2$ and generalised in [15] for any p, forms an important restriction of the theory developed in this chapter.

Definition 5.1.10 (Crooked property). A function $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ is said to have the *crooked property* if all of its differential sets $\mathrm{Im}(D_{F,a})$, for $a \in \mathbb{F}_p^n \setminus \{0\}$, are affine subspaces of $\mathbb{F}_p^m$. Furthermore, F is said to be *crooked of codimension* k when every differential set is an affine subspace of codimension k. The specific case of codimension 1, where every differential set is an affine hyperplane, is simply referred to as *crooked*.

Every quadratic function satisfies the (general) crooked property.

Proposition 5.1.11. *For a quadratic function $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ and any non-zero $a \in \mathbb{F}_p^n$, the image set $\mathrm{Im}(D_{F,a})$ is an affine subspace of $\mathbb{F}_p^m$.*

Proof. By Lemma 5.1.9, we have $D_{F,a}(x) = L_a(x) + c_a$. Since L_a is a linear map, its image $\mathrm{Im}(L_a)$ is a linear subspace of $\mathbb{F}_p^m$. Therefore, $\mathrm{Im}(D_{F,a})$ is the translation of the subspace $\mathrm{Im}(L_a)$ by the vector c_a, which by definition is an affine subspace. ∎

Interestingly, it is known that for a quadratic function, $p = 2$, and $m = n$, being crooked is equivalent to being APN.

Theorem 5.1.12. *Let $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a quadratic function. Then F is crooked if and only if F is an APN function.*

Proof. For the sufficient condition, assume F is crooked. For any $a \in \mathbb{F}_2^n \setminus \{0\}$, the differential set $\mathrm{Im}(D_{F,a})$ is an affine hyperplane; thus, its cardinality is 2^{n-1}. For any $x \in \mathbb{F}_2^n$ and $a \neq 0$, we have $D_{F,a}(x) = D_{F,a}(x + a)$. Since x and $x + a$ are distinct, every element $y \in \mathrm{Im}(D_{F,a})$ has at least two preimages. These account for at least $2 \times 2^{n-1} = 2^n$ elements. Since the domain has 2^n elements, every $y \in \mathrm{Im}(D_{F,a})$ has exactly two preimages, meaning F is APN.

Conversely, suppose F is an APN function. Because F is quadratic, its difference function $D_{F,a}(x) = F(x+a)+F(x)$ has algebraic degree at most 1 (see Lemma 5.1.9), making it an affine function. Therefore, its image $\mathrm{Im}(D_{F,a})$ must be an affine subspace of $\mathbb{F}_2^n$. Since F is APN, the equation $D_{F,a}(x) = y$ has exactly two solutions for any y in the image. Because the domain has size 2^n, the image must have $2^n/2 = 2^{n-1}$ elements. An affine subspace of size 2^{n-1} in $\mathbb{F}_2^n$ is an affine hyperplane, which by definition means F is crooked. ∎

For the case of crooked quadratic functions, we will be able to establish a link between their Walsh spectrum and vector space partitions on the one hand, and between the number of non-bent components and certain geometric structures on the other hand. In the case $p = 2$, these results are particularly interesting for cryptography due to its 1-to-1 correspondence with APN functions (if $m = n$).

5.2 Crooked quadratic functions in characteristic two

Throughout this section, we assume characteristic two, i.e. $p = 2$. For a quadratic function $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$, the Walsh spectrum is heavily restricted. The following proposition establishes the link between the Walsh transform of a component function and the radical of its associated bilinear form.

Proposition 5.2.1 ([12, Proposition 55]). *Let $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a quadratic function and let $b \in \mathbb{F}_2^m$. Let Q_b be the quadratic form associated with the component F_b (see Proposition 5.1.5), and let B_b be its associated bilinear form. Moreover, let k be the dimension of its radical, defined as*

$$\mathrm{Rad}(Q_b) = \left\{ u \in \mathbb{F}_p^n : B_b(x, u) = 0, \forall x \in \mathbb{F}_p^n \right\}. \tag{5.1}$$

Then the Walsh transform $W_F(b, a)$ takes values in $\{0, \pm 2^{\frac{n+k}{2}}\}$.

We refer to the non-zero magnitude of the Walsh transform, given by $2^{\frac{n+k}{2}}$, as the *amplitude* of F_b. Components with the smallest possible amplitude ($k = 0$) are called *bent components*.

Corollary 5.2.2. *Let $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a quadratic function and let $b \in \mathbb{F}_2^m$. Then F_b is bent if and only if B_b is non-degenerate.*

Definition 5.2.3 (Amplitude distribution). Let $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a quadratic function. We call the multiset of all amplitudes of non-trivial components of F the *amplitude distribution* of F. If F has k_i non-trivial component functions of amplitude $2^{\frac{n+i}{2}}$, then we write its amplitude distribution as $[0^{k_0}, 1^{k_1}, \ldots, n^{k_n}]$.

In the remainder of this section, we summarise the main ideas and connections of [7]. While that text focuses on quadratic APN functions $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ with n even, the required assumptions for the geometric structures we discuss will vary. We keep the restriction n even, as the amplitude distribution for n odd is completely determined: all non-trivial components have amplitude $2^{\frac{n+1}{2}}$ [13, Theorem 8]. For brevity, most proofs are omitted and the discussion regarding equivalence is excluded.

5.2.1 Vector space partitions

The connection between the Walsh spectrum and finite geometry relies on the specific structure of the difference function. For crooked quadratic (APN, see Theorem 5.1.12) functions $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$, the image set of this function is an affine hyperplane by Definition 5.1.10.

We denote the two possible hyperplanes associated with a vector $b \in \mathbb{F}_2^n \setminus \{0\}$ as

$$H_b = \left\{ x \in \mathbb{F}_2^n : \langle b, x \rangle = 0 \right\}, \quad \overline{H_b} = \left\{ x \in \mathbb{F}_2^n : \langle b, x \rangle = 1 \right\}.$$

Based on the image of the difference function, the input space is partitioned into sets associated with these hyperplanes. For a fixed $b \in \mathbb{F}_2^n \setminus \{0\}$, define

$$T_b = \left\{ a \in \mathbb{F}_2^n : \mathrm{Im}(D_{F,a}) = H_b \right\} \cup \{0\}, \quad \overline{T_b} = \left\{ a \in \mathbb{F}_2^n : \mathrm{Im}(D_{F,a}) = \overline{H_b} \right\}.$$

Combining these, we define the set $V_b = T_b \cup \overline{T_b}$. In [7], it was proven that T_b is always a subspace of $\mathbb{F}_2^n$ and that V_b is a subspace of dimension $\dim(T_b)$ or $\dim(T_b) + 1$. The collection of these spaces forms a specific structure, namely a vector space partition.

Definition 5.2.4 (Vector space partition). Let $\mathbb{F}_p^n$ be a vector space of dimension n over a finite field $\mathbb{F}_p$. A *vector space partition* (VSP) is a collection of subspaces $\mathcal{V} = \{V_1, \ldots, V_t\}$ of $\mathbb{F}_p^n$ such that every non-zero vector is contained in a unique member of $\mathcal{V}$, i.e.,

$$\mathbb{F}_p^n = \bigcup_{i=1}^{t} V_i \quad \text{and} \quad V_i \cap V_j = \{0\} \text{ for } i \neq j.$$

We say $\mathcal{V}$ is of type $[d_1^{n_1}, \ldots, d_k^{n_k}]$ if for all $i \in \{1, \ldots, k\}$, the partition $\mathcal{V}$ contains exactly n_i vector spaces of dimension d_i. For uniqueness, we set $1 \leq d_1 < d_2 < \cdots < d_k \leq n$.

Theorem 5.2.5 ([7, Theorem 2.3]). *Let $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a crooked quadratic function. The set*

$$\mathcal{V}_F = \{\, V_b : b \in \mathbb{F}_2^n \setminus \{0\}, \dim(V_b) \geq 1 \,\}$$

is a vector space partition of $\mathbb{F}_2^n$.

Crucially, the dimension of the space V_b in this partition directly gives us the amplitude of the component function F_b.

Theorem 5.2.6 ([7, Theorem 2.4]). *Let n be even, $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a crooked quadratic function, and $\mathcal{V}_F = \{\, V_b : b \in \mathbb{F}_2^n \setminus \{0\}, \dim(V_b) \geq 1 \,\}$. Then F_b has amplitude $2^{\frac{n+\dim(V_b)}{2}}$. In particular, if F has amplitude distribution $[0^{k_0}, 2^{k_2}, \ldots, n^{k_n}]$, then $\mathcal{V}_F$ is of type $[2^{k_2}, \ldots, n^{k_n}]$.*

Consequently, the amplitude distribution of F corresponds to the type of the vector space partition $\mathcal{V}_F$. But this connection also restricts the possible dimensions of V_b, because the amplitudes are necessarily integers when working over a field of characteristic two.

Corollary 5.2.7 ([7, Corollary 2.7]). *Let n be even and $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a crooked quadratic function with associated vector space partition $\mathcal{V}_F = \{\, V_b : b \in \mathbb{F}_2^n \setminus \{0\}, \dim(V_b) \geq 1 \,\}$. Then $\dim(V_b)$ is even for every $V_b \in \mathcal{V}_F$.*

While there is a direct link between the amplitude distribution of a crooked quadratic function and a corresponding vector space partition, the existence of one does not automatically guarantee the existence of the other. Specifically, it remains an open problem whether every theoretically possible amplitude distribution of a crooked quadratic function can be realised by a vector space partition (of course, if we have a concrete function, we can construct the VSP from the previous theorems). Conversely, while certain vector space partitions are known to exist, there is no guarantee they correspond to the amplitude distribution of a valid crooked quadratic function.

5.2.2 Blocking sets in projective space

Let $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a quadratic function, not necessarily crooked nor restricted to $m = n$. The study of the non-bent components of such a function F, i.e. components with amplitude $2^{\frac{n+k}{2}}$ for $1 \leq k \leq n$, provides a second perspective. Let N_F denote the set of indices of non-trivial, non-bent component functions:

$$N_F = \{\, b \in \mathbb{F}_2^m \setminus \{0\} : F_b \text{ is not bent} \,\}.$$

This set can be identified with a subset of the projective space $\mathrm{PG}(m-1, 2)$. It was shown in [7] that for any quadratic function F, the set N_F intersects every subspace $W \leq \mathbb{F}_2^m$ of dimension at least $n/2 + 1$ in an odd number of points and thus forms a blocking set with respect to $(n/2)$-spaces in $\mathrm{PG}(m-1, 2)$.

Definition 5.2.8 (Blocking set). Let $\mathrm{PG}(n, p)$ be the projective space of dimension n over $\mathbb{F}_p$. A subset $B \subseteq \mathrm{PG}(n, p)$ is called a *blocking set* with respect to k-spaces if B intersects every k-space in at least one element. If it contains an $(n-k)$-space, it is called a *trivial* blocking set. A blocking set B such that no proper subset of B is a blocking set is called *minimal*.

Proposition 5.2.9 ([7, Corollary 3.3]). *Let n be even and $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a quadratic function. Then N_F is a blocking set in $\mathrm{PG}(m-1, 2)$ with respect to $(n/2)$-spaces. Moreover, $|\mathcal{U} \cap N_F|$ is odd for each $(n/2)$-space $\mathcal{U} \subseteq \mathrm{PG}(m-1, 2)$.*

This characterisation of N_F as a blocking set allowed Bénéteau et al. [7] to establish lower bounds on its size using theorems from finite geometry, thereby also deriving upper bounds on the total number of bent components of a crooked quadratic (APN) function (for which $m = n$). However, we have succeeded in sharpening this bound by considering another geometric structure.

5.2.3 Hypersurface in projective space

Unlike the blocking set approach, viewing the set of non-bent components N_F algebraically reveals that it forms a projective hypersurface. To formalise this, we first define this geometric structure.

Definition 5.2.10 (Projective hypersurface). A projective hypersurface of degree d in $\mathrm{PG}(n-1,p)$ is the zero set of a non-zero homogeneous polynomial $P(x_1,\dots,x_n)$ of degree d; that is, the set of points $\langle x\rangle \in \mathrm{PG}(n-1,p)$ such that $P(x)=0$.

A crucial tool in the proof of the following theorem is the *Pfaffian* and its relation to the determinant of an alternating matrix, i.e. an $n\times n$ matrix A over $\mathbb{F}_p$ such that $xAx^\top = 0$ for every vector $x\in\mathbb{F}_p^n$. While the Pfaffian can be defined explicitly via permutations or perfect matchings (see, for instance, [52, pp. 588–590]), the exact combinatorial formulation is not required for our purposes. Instead, the only property we rely on is that for an $n\times n$ alternating matrix A (where n is even), its Pfaffian, denoted by $\mathrm{Pf}(A)$, is a well-defined homogeneous polynomial of degree exactly $n/2$ in the entries of A. The following lemma establishes a link between the Pfaffian and the determinant.

Lemma 5.2.11 ([14]). *Let n be even. For any $n\times n$ alternating matrix A over $\mathbb{F}_p$, the determinant of A is equal to the square of its Pfaffian:*

$$\det(A) = (\mathrm{Pf}(A))^2.$$

Next, consider the dual projective space $\mathrm{PG}(m-1,2)^*$. The points of this dual space correspond to the hyperplanes of the original output space $\mathbb{F}_2^m$. Since every such hyperplane is the kernel of a linear functional $y\mapsto\langle b,y\rangle$, the vector $b\in\mathbb{F}_2^m$ acts as the coordinate of a point in the dual space. Hence, we can identify the point $\langle b\rangle\in\mathrm{PG}(m-1,2)^*$ with the non-zero component function F_b. We can thus redefine the set N_F as a subset of this dual space:

$$N_F = \{\,\langle b\rangle \in \mathrm{PG}(m-1,2)^* : b\in\mathbb{F}_2^m, F_b \text{ is not bent}\,\}. \tag{5.2}$$

We are now ready to prove that N_F forms a projective hypersurface.

Theorem 5.2.12. *Let n be even and $F\colon\mathbb{F}_2^n\to\mathbb{F}_2^m$ be a quadratic function. The set N_F forms a projective hypersurface of degree $n/2$ in the projective space $\mathrm{PG}(m-1,2)^*$. Specifically, N_F is the zero set of a homogeneous polynomial $P(b_1,\dots,b_m)$ of degree $n/2$ with coefficients in $\mathbb{F}_2$.*

Proof. Recall from Proposition 5.1.5 and Lemma 5.1.8 that for any $b\in\mathbb{F}_2^m$, the component function F_b is associated with a symmetric bilinear form B_b represented by an $n\times n$ matrix M_b. By Corollary 5.2.2, a component function F_b is bent if and only if B_b is non-degenerate. Equivalently, F_b is bent if and only if M_b is non-singular. Therefore, the set of non-bent components is $N_F=\{\langle b\rangle\in\mathrm{PG}(m-1,2)^* : \det(M_b)=0\}$.

In characteristic two, the bilinear form B_b is alternating because $B_b(x,x)=Q_b(2x)-2Q_b(x)=0$. Therefore, M_b is an alternating matrix. In $\mathbb{F}_2$, by Lemma 5.2.11, we have

$$\det(M_b) = (\mathrm{Pf}(M_b))^2.$$

It follows that $\det(M_b)=0$ if and only if $\mathrm{Pf}(M_b)=0$.

The Pfaffian $\mathrm{Pf}(M_b)$ is a homogeneous polynomial of degree $n/2$ in the entries of the matrix M_b. By Lemma 5.1.8, each entry $(M_b)_{ij}$ is itself a homogeneous linear polynomial (degree 1) in the coordinates $b_1,\dots,b_m$. Substituting these linear forms into the Pfaffian yields a composite polynomial $P(b_1,\dots,b_m)=\mathrm{Pf}(M_b)$ that is homogeneous of degree exactly $n/2$.

In conclusion, N_F is the zero set of P in $\mathrm{PG}(m-1,2)^*$, proving that N_F is a projective hypersurface of degree $n/2$. ∎

In odd characteristic, a similar construction is attainable, albeit without the Pfaffian.

5.3 Generalisation to odd characteristics

Throughout this section, we assume $p > 2$ prime. The finite field $\mathbb{F}_{p^n}$ can be viewed as an n-dimensional vector space over $\mathbb{F}_p$. By fixing a basis of $\mathbb{F}_{p^n}$ over $\mathbb{F}_p$, we can identify the vector space $\mathbb{F}_p^n$ with the field $\mathbb{F}_{p^n}$ via a vector space isomorphism, allowing functions $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ to be represented as polynomials over $\mathbb{F}_{p^n}$. As in the previous section, we restrict ourselves to quadratic functions.

A well-studied subset of quadratic functions is the class of (pure) Dembowski-Ostrom polynomials. These were first introduced by Dembowski and Ostrom [22] to describe projective planes of order p^n with a collineation group of order p^{2n}. In this specific class of polynomials, every individual term has an algebraic degree of exactly 2.

Definition 5.3.1 (Dembowski-Ostrom polynomial). A polynomial F over $\mathbb{F}_{p^n}$ is a *Dembowski-Ostrom (DO) polynomial* if it admits a representation of the form

$$F(x) = \sum_{0 \le i,j < n} a_{ij} x^{p^i + p^j},$$

where $a_{ij} \in \mathbb{F}_{p^n}$.

We call $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ a *Dembowski-Ostrom (DO) function* if it admits the representation given in Definition 5.3.1, but we will be mainly concerned with general quadratic functions as in Definition 5.1.4.

For a quadratic function $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$, the extended Walsh spectrum takes on only a restricted set of values. We state this formally in the following proposition; a proof is provided later.

Proposition 5.3.2. *Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ be a quadratic function. For any component function F_b with $b \in \mathbb{F}_p^m$, there exists a constant $\mu_b \ge 0$ such that the magnitude of the Walsh transform satisfies*

$$|W_F(b,a)| \in \{0, \mu_b\}$$

for all $a \in \mathbb{F}_p^n$.

Since the magnitude of the Walsh transform is either zero or a constant value determined solely by the component b, the spectral properties of F_b are fully characterised by its maximum magnitude over all a.

Definition 5.3.3 (Amplitude). The *amplitude* $\mathrm{Amp}(F_b)$ of the component function F_b is defined as the maximum magnitude of its Walsh transform over all $a \in \mathbb{F}_p^n$, i.e.

$$\mathrm{Amp}(F_b) = \max_{a \in \mathbb{F}_p^n} |W_F(b,a)|.$$

Note that for a purely quadratic function where the affine part is absent, which is the case for a DO function, this maximum is obtained at $a = 0$. In such cases, the amplitude can be equivalently defined as $\mathrm{Amp}(F_b) = |W_F(b,0)|$. For a general quadratic function, however, the presence of an affine part in the component function F_b can shift this non-zero magnitude away from $a = 0$, which is why we must take the maximum over all $a \in \mathbb{F}_p^n$ to obtain the non-zero magnitude.

This new definition aligns with the case $p = 2$. As established earlier, the Walsh spectrum of a quadratic function in characteristic two is also two-valued – taking only zero or a single non-zero magnitude – and for that case too it is the non-zero value that is defined as the amplitude.

5.3.1 Vector space partitions

Throughout this discussion on vector space partitions, **let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a crooked quadratic function (for which $m = n$)**. Since the differential set $\mathrm{Im}(D_{F,a})$ is a translation of the image of L_a, its

codimension matches the dimension of the kernel of L_a. Hence, being crooked is equivalent to imposing the following restriction on the linearised part L_a:

$$\dim(\ker(L_a)) = 1, \quad \text{for all } a \in \mathbb{F}_p^n \setminus \{0\}. \tag{5.3}$$

Indeed, the rank-nullity theorem gives $\dim(\mathrm{Im}(D_{F,a})) = n - 1$.

Furthermore, since $\dim(\ker(L_a)) = 1$, the kernel contains exactly p elements. Therefore, for every $b \in \mathbb{F}_p^n$ and $a \in \mathbb{F}_p^n \setminus \{0\}$, the equation $F(x+a) - F(x) = b$ has exactly p solutions. In particular, this implies that the crooked quadratic function F has differential uniformity p.

Example 5.3.4. Consider the vector space $\mathbb{F}_p^n$ with n even. Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be the DO function defined by the Gold monomial [36]:

$$F(x) = x^{p+1}.$$

For a non-zero $a \in \mathbb{F}_p^n$, we compute the difference function $D_{F,a}$ using the decomposition established in Lemma 5.1.9:

$$\begin{aligned} D_{F,a}(x) &= (x+a)^{p+1} - x^{p+1} \\ &= \underbrace{ax^p + a^p x}_{L_a(x)} + a^{p+1}. \end{aligned}$$

To determine the kernel of the linearised map L_a, we solve $ax^p + a^p x = 0$. Substituting $x = az$, and dividing by the non-zero factor a^{p+1}, we obtain the equivalent equation

$$z^p + z = 0.$$

The polynomial $z^p + z$ corresponds to the trace function $\mathrm{Tr}_{\mathbb{F}_{p^2}/\mathbb{F}_p}(z)$. Since n is even, $\mathbb{F}_{p^2}$ is a subfield of $\mathbb{F}_{p^n}$. The kernel of this trace map has dimension 1 over $\mathbb{F}_p$, consisting of exactly p elements. Since a polynomial of degree p can have at most p roots, these kernel elements constitute the entire solution set. Thus, all roots lie within $\mathbb{F}_{p^2} \subseteq \mathbb{F}_{p^n}$.

So, $|\ker(L_a)| = p$ for all $a \in \mathbb{F}_p^n \setminus \{0\}$, and the image of $D_{F,a}$ is an affine hyperplane. Thus, the Gold monomial x^{p+1} satisfies the imposed restriction (5.3) when n is even and is therefore crooked. □

The affine hyperplanes we obtain by imposing restriction (5.3), i.e. imposing that the quadratic function is crooked, will be denoted by, for $b \in \mathbb{F}_p^n \setminus \{0\}$ and $\gamma \in \mathbb{F}_p$,

$$H_{b,\gamma} := \left\{ x \in \mathbb{F}_p^n : \mathrm{Tr}_{\mathbb{F}_p}(bx) = \gamma \right\}.$$

This covers all possibilities, because any affine hyperplane is defined as the level set of a non-zero linear functional $\phi\colon \mathbb{F}_p^n \to \mathbb{F}_p$, and every such functional can be uniquely represented as $x \mapsto \mathrm{Tr}_{\mathbb{F}_p}(bx)$ for some $b \in \mathbb{F}_p^n \setminus \{0\}$ [55, Theorem 2.24].

Following the approach for $p = 2$, we further define

$$T_{b,0} := \left\{ a \in \mathbb{F}_p^n : \mathrm{Im}(D_{F,a}) = H_{b,0} \right\} \cup \{0\} \quad \text{and} \quad T_{b,\gamma} := \left\{ a \in \mathbb{F}_p^n : \mathrm{Im}(D_{F,a}) = H_{b,\gamma} \right\},$$

for $\gamma \in \mathbb{F}_p^*$.

Proposition 5.3.5. *For any $b \in \mathbb{F}_p^n \setminus \{0\}$, the following hold:*

(i) $T_{b,0}$ is a vector space over $\mathbb{F}_p$, in particular a subspace of $\mathbb{F}_p^n$.

(ii) If $T_{b,\gamma}$ is non-empty for some $\gamma \in \mathbb{F}_p^*$, it is an affine space over $\mathbb{F}_p$, specifically a translation of $T_{b,0}$.

(iii) $V_b := \bigcup_{\gamma \in \mathbb{F}_p} T_{b,\gamma}$ is a vector space over $\mathbb{F}_p$.

Proof. (i) By definition, $0 \in T_{b,0}$. Let $a_1, a_2 \in T_{b,0}$. For any $x \in \mathbb{F}_p^n$, we observe

$$F(x + a_1 + a_2) - F(x) = (F(x + a_1 + a_2) - F(x + a_1)) + (F(x + a_1) - F(x)) .$$

Let $y = x + a_1$. Since $a_2 \in T_{b,0}$, the term $F(y + a_2) - F(y)$ lies in $H_{b,0}$. Similarly, since $a_1 \in T_{b,0}$, the term $F(x + a_1) - F(x)$ lies in $H_{b,0}$. As $H_{b,0}$ is a linear subspace (the kernel of the trace function), it is closed under addition. So, $F(x + a_1 + a_2) - F(x) \in H_{b,0}$ for all $x \in \mathbb{F}_p^n$, which implies $a_1 + a_2 \in T_{b,0}$. Thus, $T_{b,0}$ is a subspace of $\mathbb{F}_p^n$.

(ii) Let $\gamma \in \mathbb{F}_p^*$ and suppose $c \in T_{b,\gamma}$. To prove that $T_{b,\gamma}$ is an affine space, we show that $T_{b,\gamma} = c + T_{b,0}$. Let $a \in T_{b,0}$. Using the same additivity argument as above, we find

$$F(x + c + a) - F(x) = (F(x + c + a) - F(x + a)) + (F(x + a) - F(x)) .$$

The first term lies in $H_{b,\gamma}$ (since $c \in T_{b,\gamma}$) and the second in $H_{b,0}$ (since $a \in T_{b,0}$). Their sum lies in $H_{b,\gamma}$, implying $c + a \in T_{b,\gamma}$. Thus, $c + T_{b,0} \subseteq T_{b,\gamma}$.

Conversely, suppose $a' \in T_{b,\gamma}$. We want to show $a' - c \in T_{b,0}$. We observe that

$$F(x + a') - F(x) = \big(F(x + a') - F(x + a' - c)\big) + \big(F(x + a' - c) - F(x)\big) .$$

Rearranging this yields

$$F(x + a' - c) - F(x) = \big(F(x + a') - F(x)\big) - \big(F(x + a') - F(x + a' - c)\big) .$$

The first term on the right is in $H_{b,\gamma}$ because $a' \in T_{b,\gamma}$. Substituting $y = x + a' - c$ into the second term gives $F(y + c) - F(y)$, which is also in $H_{b,\gamma}$ because $c \in T_{b,\gamma}$. The difference of two elements in the affine hyperplane $H_{b,\gamma}$ lies in the linear hyperplane $H_{b,0}$. Thus, $a' - c \in T_{b,0}$, meaning $T_{b,\gamma} \subseteq c + T_{b,0}$.

(iii) We must show that V_b is closed under linear combinations over $\mathbb{F}_p$. By the same logic as above, if $a_1 \in T_{b,\gamma_1}$ and $a_2 \in T_{b,\gamma_2}$, their sum $a_1 + a_2 \in T_{b,\gamma_1+\gamma_2} \subseteq V_b$. It remains to verify closure under scalar multiplication. Let $a \in T_{b,\gamma}$ and let $\lambda \in \mathbb{F}_p$. We can write the difference as a telescoping sum:

$$F(x + \lambda a) - F(x) = \sum_{i=1}^{\lambda} (F(x + ia) - F(x + (i - 1)a)) .$$

Each term in the sum is of the form $F(y + a) - F(y)$ for $y = x + (i - 1)a$. Since $a \in T_{b,\gamma}$, every term belongs to $H_{b,\gamma}$. The sum of λ elements from $H_{b,\gamma}$ lies in $H_{b,\lambda\gamma}$. Therefore, $\lambda a \in T_{b,\lambda\gamma} \subseteq V_b$. We conclude that V_b is a vector space over $\mathbb{F}_p$. ∎

For the set V_b, we can also adopt an equivalent definition based on the theory of quadratic forms. Let Q_b be the quadratic form associated with the component F_b (as established in Proposition 5.1.5), and let B_b denote its associated symmetric bilinear form. The following results show the correspondence between $\mathrm{Rad}(Q_b)$ (see Eq. (5.1)) and the space V_b, mirroring the results known for $p = 2$.

Lemma 5.3.6. *For any $b \in \mathbb{F}_p^n \setminus \{0\}$, it holds that $a \in \mathrm{Rad}(Q_b)$ if and only if $D_{Q_b,a}(x) = 0$ for all $x \in \mathbb{F}_p^n$.*

Proof. Observe that $D_{Q_b,a}(x) = Q_b(x + a) - Q_b(x) = B_b(x, a) + Q_b(a)$.

First, suppose $a \in \mathrm{Rad}(Q_b)$. By definition, $B_b(x, a) = 0$ for all $x \in \mathbb{F}_p^n$. Furthermore, since the characteristic p is odd, we have $Q_b(a) = 2^{-1}B_b(a, a) = 0$. Substituting these into the difference function yields $D_{Q_b,a}(x) = 0$ for all x.

Conversely, assume $D_{Q_b,a}(x) = 0$ for all $x \in \mathbb{F}_p^n$. Then $B_b(x, a) = -Q_b(a)$ for all x. The left-hand side, $x \mapsto B_b(x, a)$, defines a linear functional on $\mathbb{F}_p^n$. The right-hand side is a constant independent of x. A linear map is constant if and only if it is the zero map (since it must map 0 to 0). Therefore, we must have $-Q_b(a) = 0$ and $B_b(x, a) = 0$ for all $x \in \mathbb{F}_p^n$. Hence, $a \in \mathrm{Rad}(Q_b)$. ∎

Proposition 5.3.7. *For every $b \in \mathbb{F}_p^n \setminus \{0\}$, it holds that $V_b = \mathrm{Rad}(Q_b)$.*

Proof. From Proposition 5.1.5 we know that the component function $F_b(x) = \mathrm{Tr}_{\mathbb{F}_p}(bF(x))$ decomposes into a quadratic form Q_b and an affine function L_b. Therefore, the difference function of the component function is

$$\mathrm{Tr}_{\mathbb{F}_p}(bD_{F,a}(x)) = D_{F_b,a}(x) = D_{Q_b,a}(x) + L_b(a),$$

with $L_b(a) = D_{L_b,a}(x)$ a constant.

By definition, $a \in V_b = \bigcup_{\gamma \in \mathbb{F}_p} T_{b,\gamma}$ if and only if the image of $D_{F,a}$ is contained in an affine hyperplane $H_{b,\gamma}$ for some $\gamma \in \mathbb{F}_p$. This is equivalent to saying $\mathrm{Tr}_{\mathbb{F}_p}(bD_{F,a}(x)) = \gamma$ for all $x \in \mathbb{F}_p^n$. Substituting our expansion, we have $D_{Q_b,a}(x) + L_b(a) = \gamma$ for all x.

Since $L_b(a)$ and γ are both constants independent of x, this holds if and only if $D_{Q_b,a}(x)$ is a constant for all $x \in \mathbb{F}_p^n$. As established in the proof of Lemma 5.3.6, $D_{Q_b,a}(x) = B_b(x,a) + Q_b(a)$. For this expression to be constant, the linear map $x \mapsto B_b(x,a)$ must be identically zero. Because the characteristic is odd, this implies $Q_b(a) = 2^{-1}B_b(a,a) = 0$. Thus, $D_{Q_b,a}(x)$ is a constant if and only if it is identically zero. Finally, by Lemma 5.3.6, $D_{Q_b,a}(x) = 0$ for all $x \in \mathbb{F}_p^n$ is a necessary and sufficient condition for $a \in \mathrm{Rad}(Q_b)$. Therefore, $V_b = \mathrm{Rad}(Q_b)$. ∎

Remark 5.3.8. Intuitively, the equivalence $V_b = \mathrm{Rad}(Q_b)$ highlights that the affine part L_b of the component function has no impact on the space V_b. Because the difference function acts as a discrete derivative, the affine part reduces to a constant $L_b(a)$. Geometrically, this constant shifts the image of the difference function from one parallel hyperplane $H_{b,\gamma}$ to another. Since V_b is defined as the union over all possible shifts $\gamma \in \mathbb{F}_p$, this affine contribution is absorbed, leaving the structure of V_b to be determined solely by the quadratic form Q_b.

Corollary 5.3.9. *V_b is a subspace of $\mathbb{F}_p^n$.*

Analogously to the binary case, we establish that the sets V_b, for $b \in \mathbb{F}_p^n \setminus \{0\}$, constitute a vector space partition of $\mathbb{F}_p^n$ as in Definition 5.2.4. However, we must account for scalar multiplicity, which is non-trivial in odd characteristic. To ensure the subspaces in our collection are distinct, we index them by one-dimensional subspaces rather than individual vectors. We denote the span of a vector b by $\langle b \rangle = \{\, \lambda b : \lambda \in \mathbb{F}_p \,\}$.

Theorem 5.3.10. *The set $\left\{\, V_{\langle b \rangle} : b \in \mathbb{F}_p^n \setminus \{0\}, \dim(V_{\langle b \rangle}) \geq 1 \,\right\}$ is a vector space partition of $\mathbb{F}_p^n$.*

Proof. Let b_1 and b_2 be two linearly independent vectors in $\mathbb{F}_p^n$. Suppose there exists a non-zero vector $u \in V_{b_1} \cap V_{b_2}$. By definition, $u \in V_{b_1}$ implies $u \in T_{b_1,\gamma_1}$ for some $\gamma_1 \in \mathbb{F}_p$, meaning the image $\mathrm{Im}(D_{F,u})$ is an affine hyperplane with normal vector b_1. Similarly, since $u \in V_{b_2}$, $\mathrm{Im}(D_{F,u})$ is an affine hyperplane with normal vector b_2. Because an affine hyperplane uniquely determines its normal vector up to a scalar multiple, it follows that $b_1 = \lambda b_2$ for some $\lambda \in \mathbb{F}_p$. This contradicts the assumption that b_1 and b_2 are linearly independent.

Now, let $a \in \mathbb{F}_p^n \setminus \{0\}$. As F is assumed crooked in this section, the image $\mathrm{Im}(D_{F,a})$ is an affine hyperplane. Every affine hyperplane in $\mathbb{F}_p^n$ is defined by a normal vector $b \in \mathbb{F}_p^n \setminus \{0\}$ and a shift $\gamma \in \mathbb{F}_p$, such that $\mathrm{Im}(D_{F,a}) = H_{b,\gamma}$. This implies $a \in T_{b,\gamma}$. By definition, it immediately follows that $a \in \bigcup_{\gamma \in \mathbb{F}_p} T_{b,\gamma} = V_b$. ∎

We can thus associate to every crooked quadratic function F for which $m = n$ a uniquely defined vector space partition $\mathcal{V}_F = \left\{\, V_{\langle b \rangle} : b \in \mathbb{F}_p^n \setminus \{0\}, \dim(V_{\langle b \rangle}) \geq 1 \,\right\}$. Note that the vector space partition is indexed by one-dimensional subspaces $\langle b \rangle$, which differs from the case of characteristic two.

Before linking the amplitude of the component functions to the spaces in this partition, we prove that the amplitude of a component function F_b is well-defined, as stated earlier under more relaxed assumptions (general quadratic function $F \colon \mathbb{F}_p^n \to \mathbb{F}_p^m$).

Proof of Proposition 5.3.2. Fix $b \in \mathbb{F}_p^m$ and consider the Walsh transform at $a \in \mathbb{F}_p^n$. By Proposition 5.1.5, the component function decomposes as $F_b(x) = Q_b(x) + L_b(x)$, where Q_b is a quadratic form and L_b is an affine function. Now note that any affine function from $\mathbb{F}_p^n$ to $\mathbb{F}_p$ can be written (uniquely) as $L_b(x) = \mathrm{Tr}_{\mathbb{F}_p}(l_b x) + \gamma_b$ for some vector $l_b \in \mathbb{F}_p^n$ and constant $\gamma_b \in \mathbb{F}_p$. Substituting this into the Walsh sum yields

$$W_F(b,a) = \sum_{x \in \mathbb{F}_p^n} \xi_p^{Q_b(x) + \mathrm{Tr}_{\mathbb{F}_p}(l_b x) + \gamma_b - \mathrm{Tr}_{\mathbb{F}_p}(ax)} = \xi_p^{\gamma_b} \sum_{x \in \mathbb{F}_p^n} \xi_p^{Q_b(x) + \mathrm{Tr}_{\mathbb{F}_p}((l_b - a)x)}.$$

Let $a' = l_b - a$. We attempt to complete the square by finding a shift $c \in \mathbb{F}_p^n$ that cancels the linear terms in the exponent. Expanding the quadratic form at $x + c$, we find $Q_b(x+c) = Q_b(x) + B_b(x,c) + Q_b(c)$. Substituting $Q_b(x) = Q_b(x+c) - B_b(x,c) - Q_b(c)$ into the sum, we wish to eliminate the linear part $\mathrm{Tr}_{\mathbb{F}_p}(a'x) - B_b(x,c)$ by choosing c such that

$$B_b(x,c) = \mathrm{Tr}_{\mathbb{F}_p}(a'x), \quad \forall x \in \mathbb{F}_p^n. \tag{5.4}$$

Consider the linear functional $\Phi_b \colon c \mapsto B_b(\cdot, c)$. The kernel of Φ_b is equal to $\mathrm{Rad}(Q_b)$. By the rank-nullity theorem, the image of Φ_b has dimension $n - \dim(\mathrm{Rad}(Q_b))$. Observe that for any c, the functional $B_b(\cdot, c)$ vanishes on $\mathrm{Rad}(Q_b)$, since $B_b(u,c) = 0$ for all $u \in \mathrm{Rad}(Q_b)$. Therefore, the image of Φ_b is contained in the annihilator of $\mathrm{Rad}(Q_b)$. But because the annihilator of $\mathrm{Rad}(Q_b)$ also has dimension $n - \dim(\mathrm{Rad}(Q_b))$, the image of Φ_b must coincide with this annihilator. This implies that Equation (5.4) admits a solution if and only if the target functional $x \mapsto \mathrm{Tr}_{\mathbb{F}_p}(a'x)$ belongs to the annihilator, meaning $\mathrm{Tr}_{\mathbb{F}_p}(a'u) = 0$ for all $u \in \mathrm{Rad}(Q_b)$.

First, assume such a solution c exists. The linear terms cancel by substituting $\mathrm{Tr}_{\mathbb{F}_p}(a'x) = B_b(x,c)$, and the exponent simplifies to $Q_b(x+c) - Q_b(c)$. Changing the summation variable to $y = x + c$, we obtain

$$W_F(b,a) = \xi_p^{\gamma_b - Q_b(c)} \sum_{y \in \mathbb{F}_p^n} \xi_p^{Q_b(y)}.$$

Taking the magnitude, the phase factor vanishes, which yields $|W_F(b,a)| = |\sum_{y \in \mathbb{F}_p^n} \xi_p^{Q_b(y)}|$. We define this constant magnitude as μ_b and observe it is only dependent on Q_b.

Conversely, suppose no solution exists. This occurs precisely when the functional $x \mapsto \mathrm{Tr}_{\mathbb{F}_p}(a'x)$ does not vanish on $\mathrm{Rad}(Q_b)$. Hence, there exists a vector $u_0 \in \mathrm{Rad}(Q_b)$ such that $\mathrm{Tr}_{\mathbb{F}_p}(a'u_0) \neq 0$. Note that for any $u_0 \in \mathrm{Rad}(Q_b)$, we have $B_b(x, u_0) = 0$ for all x, and thus $Q_b(u_0) = 2^{-1} B_b(u_0, u_0) = 0$. Applying the shift $x \mapsto x + u_0$ to the sum yields

$$\begin{aligned}\sum_{x \in \mathbb{F}_p^n} \xi_p^{Q_b(x+u_0) + \mathrm{Tr}_{\mathbb{F}_p}(a'(x+u_0))} &= \sum_{x \in \mathbb{F}_p^n} \xi_p^{Q_b(x) + B_b(x,u_0) + Q_b(u_0) + \mathrm{Tr}_{\mathbb{F}_p}(a'x) + \mathrm{Tr}_{\mathbb{F}_p}(a'u_0)} \\ &= \xi_p^{\mathrm{Tr}_{\mathbb{F}_p}(a'u_0)} \sum_{x \in \mathbb{F}_p^n} \xi_p^{Q_b(x) + \mathrm{Tr}_{\mathbb{F}_p}(a'x)}.\end{aligned}$$

Since $\mathrm{Tr}_{\mathbb{F}_p}(a'u_0) \neq 0$, we have $\xi_p^{\mathrm{Tr}_{\mathbb{F}_p}(a'u_0)} \neq 1$, which implies the sum $\sum_x \xi_p^{Q_b(x) + \mathrm{Tr}_{\mathbb{F}_p}(a'x)}$ evaluates to 0. This relies on the fact that shifting the summation variable by u_0 leaves the total sum invariant, equating the original sum to its phase-shifted counterpart. In this case, we conclude that $W_F(b,a) = 0$. ∎

Analogously to Theorem 5.2.6, for odd characteristic, the dimension of a subspace $V_{\langle b \rangle} \in \mathcal{V}_F$ is also directly linked to the amplitude of F_b.

Theorem 5.3.11. *Let $F \colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a crooked quadratic function, and let $\mathcal{V}_F$ be its associated vector space partition. Then, for every $b \in \mathbb{F}_p^n \setminus \{0\}$, the component function F_b has amplitude $p^{\frac{n + \dim(V_{\langle b \rangle})}{2}}$.*

Proof. Fix $b \in \mathbb{F}_p^n \setminus \{0\}$. Recall from the proof of Proposition 5.3.2 that the amplitude of F_b is exactly the constant magnitude $\mu_b = |\sum_{x \in \mathbb{F}_p^n} \xi_p^{Q_b(x)}|$, where Q_b is the quadratic form associated with F_b. Let B_b be the symmetric bilinear form associated with Q_b. By definition, the subspace $V_b = \mathrm{Rad}(Q_b)$ is the kernel of the linear map defined by B_b. Note that V_b corresponds precisely to the subspace $V_{\langle b \rangle}$ in the partition $\mathcal{V}_F$ (since $V_b = V_{\lambda b}$ for any $\lambda \in \mathbb{F}_p^*$). If we denote the rank of the quadratic form Q_b by r, the rank-nullity theorem yields

$$\dim(V_{\langle b \rangle}) = \dim(V_b) = n - r.$$

As the characteristic p is odd, there exists a coordinate transformation $\phi \colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ that diagonalises the quadratic form [55, Theorem 6.21]. Let $x \in \mathbb{F}_p^n$ and let $y = \phi(x) = (y_1, \ldots, y_n)$ be the coordinate vector of x in the new basis. In this coordinate system, the quadratic form can be written as

$$Q_b(x) = \sum_{i=1}^{r} \lambda_i y_i^2,$$

where $\lambda_i \in \mathbb{F}_p^*$ for $1 \leq i \leq r$. The variables $y_{r+1}, \ldots, y_n$ do not appear in this sum as they correspond to the radical V_b.

We now compute the sum defining μ_b using this coordinate change. Since ϕ is a bijection, summing over $x \in \mathbb{F}_p^n$ is equivalent to summing over $y \in \mathbb{F}_p^n$. We can factor the sum as follows:

$$\begin{aligned}
\sum_{x \in \mathbb{F}_p^n} \xi_p^{Q_b(x)} &= \sum_{y_1, \ldots, y_n \in \mathbb{F}_p} \xi_p^{\lambda_1 y_1^2 + \cdots + \lambda_r y_r^2} \\
&= \sum_{y_1, \ldots, y_n \in \mathbb{F}_p} \left(\prod_{i=1}^{r} \xi_p^{\lambda_i y_i^2} \right) \\
&= \left(\sum_{y_1 \in \mathbb{F}_p} \xi_p^{\lambda_1 y_1^2} \right) \cdots \left(\sum_{y_r \in \mathbb{F}_p} \xi_p^{\lambda_r y_r^2} \right) \left(\sum_{y_{r+1} \in \mathbb{F}_p} 1 \right) \cdots \left(\sum_{y_n \in \mathbb{F}_p} 1 \right).
\end{aligned}$$

This expression contains two types of factors. The last $n - r$ factors are simply sums of 1 over $\mathbb{F}_p$, each contributing a factor of p. The first r factors are standard quadratic Gauss sums. Now, for any $c \in \mathbb{F}_p^*$, the magnitude of the Gauss sum is given by [55, Theorem 5.11]:

$$|\sum_{z \in \mathbb{F}_p} \xi_p^{cz^2}| = \sqrt{p}.$$

Taking the magnitude of our factored sum and substituting these known values, we obtain the amplitude

$$\mathrm{Amp}(F_b) = \mu_b = \left(\prod_{i=1}^{r} \sqrt{p} \right) \cdot p^{n-r} = p^{r/2} \cdot p^{n-r} = p^{n - r/2}.$$

Finally, substituting $r = n - \dim\left(V_{\langle b \rangle}\right)$, we conclude that

$$\mathrm{Amp}(F_b) = p^{n - \frac{n - \dim\left(V_{\langle b \rangle}\right)}{2}} = p^{\frac{n + \dim\left(V_{\langle b \rangle}\right)}{2}}.$$ ∎

From Theorem 5.3.11, we know that the amplitude of F_b is determined solely by the dimension of $V_{\langle b \rangle}$. This dependence leads to a natural grouping of component functions.

Remark 5.3.12. Since $V_{\lambda b} = V_b$ for any non-zero scalar $\lambda \in \mathbb{F}_p$, the amplitude is invariant on the one-dimensional subspace $\langle b \rangle$. So, each element $V \in \mathcal{V}_F$ corresponds to exactly $p - 1$ component functions sharing the same amplitude.

Motivated by this observation, we define the amplitude distribution for a crooked quadratic function in a way that factors out this scalar multiplicity.

Definition 5.3.13 (Amplitude distribution). Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a crooked quadratic function. We call the multiset of all amplitudes of non-trivial components of F the *amplitude distribution* of F. If F has $(p-1)k_i$ non-trivial component functions of amplitude $p^{\frac{n+i}{2}}$, then we write its amplitude distribution as $[0^{(p-1)k_0}, 1^{(p-1)k_1}, \ldots, n^{(p-1)k_n}]$.

This notation aligns with the structure highlighted in Remark 5.3.12, where the exponent k_i counts the number of subspaces of dimension i in the partition rather than the raw count of component functions.

5.3.2 Hypersurface in projective space

Consider a quadratic function $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$, not necessarily crooked nor restricted to $m = n$. Unlike in the binary case, for finite fields with characteristic $p > 2$, the set

$$N_F = \left\{ b \in \mathbb{F}_p^m \setminus \{0\} : F_b \text{ is not bent} \right\}$$

is not necessarily a blocking set. But, just as we established for the binary case (see Eq. (5.2)), we can identify its non-zero component functions F_b (up to scalar multiplication) with points $\langle b \rangle$ in the dual projective space $\mathrm{PG}(m-1,p)^*$.

For any quadratic function, a component F_b is bent if and only if its associated symmetric bilinear form B_b is non-degenerate. This is equivalent to its radical $\mathrm{Rad}(Q_b)$ being trivial. Therefore, F_b is non-bent if and only if $\dim(\mathrm{Rad}(Q_b)) > 0$. We can thus redefine the set N_F of non-bent components in this geometric setting as

$$N_F = \left\{ \langle b \rangle \in \mathrm{PG}(m-1,p)^* : b \in \mathbb{F}_p^m \setminus \{0\}, \dim(\mathrm{Rad}(Q_b)) > 0 \right\}. \tag{5.5}$$

Note that if the quadratic function F is crooked (and $m = n$), Proposition 5.3.7 ensures that $\mathrm{Rad}(Q_b)$ coincides exactly with the subspace V_b of the associated vector space partition.

A counterexample showing that N_F is not necessarily a blocking set is provided next.

Example 5.3.14. Let $n = 2$, $m = 3$, and $p > 2$. Consider the quadratic function $F\colon \mathbb{F}_p^2 \to \mathbb{F}_p^3$ defined by $F(x,y) = (x^2, xy, y^2)$. While not strictly necessary to construct a counterexample (we must only assume that F is quadratic), we first verify that this function is crooked. For any non-zero $a = (a_1, a_2) \in \mathbb{F}_p^2$, evaluating the difference function $D_{F,a}$ reveals its linear part L_a to be

$$L_a(x,y) = (2a_1x, a_2x + a_1y, 2a_2y).$$

Since p is odd, the kernel of L_a requires $2a_1x = 0$ and $2a_2y = 0$. Because a is non-zero, at least one of its coordinates is non-zero, which forces $x = y = 0$. Thus, $\ker(L_a)$ is trivial, and the image $\mathrm{Im}(L_a)$ has dimension 2. In $\mathbb{F}_p^3$, an affine subspace of dimension 2 is a hyperplane. Therefore, F is a crooked quadratic function.

We now identify the set of non-bent components N_F in the dual projective space $\mathrm{PG}(2,p)^*$; see Equation (5.5). Let $b = (b_1, b_2, b_3) \in \mathbb{F}_p^3$. The component function expands to $F_b(x,y) = \langle b, F(x,y) \rangle = b_1x^2 + b_2xy + b_3y^2$. As introduced in Lemma 5.1.8, the symmetric matrix M_b representing the associated bilinear form is

$$M_b = \begin{pmatrix} 2b_1 & b_2 \\ b_2 & 2b_3 \end{pmatrix}.$$

As established, a component F_b is non-bent precisely when $\dim(\mathrm{Rad}(Q_b)) > 0$, which is equivalent to the matrix M_b being singular, i.e. $\det(M_b) = 0$. Calculating the determinant yields the following equation:

$$4b_1b_3 - b_2^2 = 0.$$

We can determine the number of non-zero vectors (b_1, b_2, b_3) satisfying the equation $4b_1b_3 = b_2^2$ to obtain the number of projective points on this (non-degenerate) conic. If $b_1 \neq 0$ ($p-1$ choices), then for any of the p choices for b_2, the coordinate b_3 is uniquely determined as $(4b_1)^{-1}b_2^2$, yielding $p(p-1)$ vectors. If $b_1 = 0$, the equation forces $b_2 = 0$, leaving $p-1$ non-zero choices for b_3. This yields another $p-1$ vectors. In total, there are $(p+1)(p-1)$ non-zero vectors. Dividing by $p-1$ to account for scalar equivalence, we find the that conic contains exactly $p+1$ points in $\mathrm{PG}(2,p)^*$.

For N_F to be a blocking set with respect to $(n/2)$-spaces, it must intersect every projective line within $\mathrm{PG}(2,p)^*$. Note that $\mathrm{PG}(2,p)^*$ contains exactly p^2+p+1 lines (through duality with the number of points in $\mathrm{PG}(2,p)$). At each point of the conic, there is exactly one line that intersects it in 1 point, the tangent line. Any other line passing through the conic intersects it in 2 points.

By counting the incidences between the $p+1$ points and the lines containing them, the total number of lines intersecting the conic is the $p+1$ tangents plus $\frac{1}{2}p(p+1)$ lines intersecting in two points, yielding a total of $\frac{1}{2}(p+1)(p+2)$ intersected lines. For any $p > 2$, this maximum number of intersected lines is strictly less than p^2+p+1. Thus, there exist exterior lines in $\mathrm{PG}(2,p)^*$ that completely miss the conic N_F.

While our set N_F is constructed in the dual space $\mathrm{PG}(m-1,p)^*$, a projective space and its dual are isomorphic as incidence structures. Because there exists a line, which corresponds to an $(n/2)$-space for $n = 2$, that completely misses N_F in the dual space, and incidence properties are preserved under isomorphism, it directly follows that the corresponding point set cannot form a blocking set in the original space either. □

However, it does retain the algebraic structure of a projective hypersurface that we observed for the case $p = 2$.

Lemma 5.3.15. *Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ be a quadratic function. The set N_F forms a projective hypersurface of degree n in the dual space $\mathrm{PG}(m-1,p)^*$. Specifically, N_F is the zero set of a homogeneous polynomial $P(b_1,\ldots,b_m)$ of degree n with coefficients in $\mathbb{F}_p$.*

Proof. For any $b \in \mathbb{F}_p^m$, let M_b be the $n \times n$ matrix representing the symmetric bilinear form B_b associated with the component function F_b. By definition, $\langle b \rangle \in N_F$ if and only if $\dim(\mathrm{Rad}(Q_b)) > 0$. Since $\mathrm{Rad}(Q_b)$ is exactly the kernel of the linear map represented by M_b (see Equation (5.1)), this condition is equivalent to M_b being singular, i.e. $\det(M_b) = 0$.

By Lemma 5.1.8, each entry $(M_b)_{ij}$ is a homogeneous linear polynomial (degree 1) in the coordinates $b_1,\ldots,b_m$. Therefore, its determinant is a function of these coordinates; let $P(b_1,\ldots,b_m) = \det(M_b)$. By the Leibniz formula for determinants,

$$\det(M_b) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} (M_b)_{i,\sigma(i)}.$$

Every term in this sum consists of a product of exactly n matrix entries, which are homogeneous linear polynomials. When n such entries are multiplied together, their degrees sum to n, making each product a homogeneous polynomial of degree n. Since P is a sum of these degree-n polynomials, it is itself a homogeneous polynomial of total degree n.

Thus, N_F is precisely the set of points in $\mathrm{PG}(m-1,p)^*$ where the homogeneous polynomial P of degree n vanishes. ■

Example 5.3.16. We revisit the case of $n = 2, m = 3$ and the function $F(x,y) = (x^2, xy, y^2)$ discussed previously (see Example 5.3.14). We established that the condition for a component $b = (b_1, b_2, b_3)$ to be in N_F is $4b_1b_3 - b_2^2 = 0$.

We now verify that this aligns perfectly with Lemma 5.3.15. The polynomial defining the hypersurface is explicitly

$$P(b_1, b_2, b_3) = 4b_1b_3 - b_2^2.$$

As predicted, P is a homogeneous polynomial of degree exactly $n = 2$. Geometrically, the hypersurface N_F in $\mathrm{PG}(2, p)^*$ is a conic section. While it failed to be a blocking set, it perfectly adheres to the structure of a projective hypersurface of degree n. □

By viewing N_F as a projective hypersurface in the dual space rather than a blocking set, we will still be able to deduce specific structural properties, particularly when $p > n$.

5.4 Conditions obtained through novel connections

The structural links established in the preceding sections provide a framework for deriving constraints on the achievable properties of vectorial functions with respect to their Walsh spectrum or non-bent components. From this, we can deduce necessary conditions for the existence of functions with specific cryptographic parameters.

In this section, we systematically exploit these connections to bound the properties of (crooked) quadratic functions (which are APN in characteristic two when $m = n$, see Theorem 5.1.12). We begin by establishing fundamental restrictions on the amplitude distribution, derived from the combinatorial limits of vector space partitions and complemented by other theoretical results from this domain. Subsequently, we focus on the set of non-bent components N_F. We derive bounds on its cardinality based on the partition type and finally explore structural constraints arising from its specific geometric nature.

5.4.1 Conditions on the amplitude distribution

Fundamental conditions follow from counting arguments and the dimension formula. The following two theorems generalise the results originally established for the binary case in [7, Theorems 2.8 and 2.10], extending the proofs to accommodate crooked quadratic functions in general characteristic p.

Theorem 5.4.1. *Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a crooked quadratic function, with amplitude distribution $[0^{k_0}, 2^{k_2}, \ldots, n^{k_n}]$ for $p = 2$ and n even, or $[0^{(p-1)k_0}, 1^{(p-1)k_1}, \ldots, n^{(p-1)k_n}]$ for $p > 2$.*

(i) $\sum_{i=1}^{n} k_i(p^i - 1) = p^n - 1$ *(packaging bound).*

(ii) If $i > \frac{n}{2}$, then $k_i \leq 1$.

(iii) If $i \neq j$ and $k_i, k_j \geq 1$, then $i + j \leq n$.

Proof. In both cases, the multiplicity k_i counts exactly the number of subspaces of dimension i in the associated vector space partition $\mathcal{V}_F$ (by Theorem 5.2.6 for $p = 2$, and Theorem 5.3.11 and Definition 5.3.13 for $p > 2$). Since $\mathcal{V}_F$ partitions the non-zero vectors of $\mathbb{F}_p^n$, summing the cardinalities yields

$$\sum_{i=1}^{n} k_i(p^i - 1) = |\mathbb{F}_p^n \setminus \{0\}| = p^n - 1,$$

proving (i). For (ii) and (iii), consider any two distinct subspaces $U, W \in \mathcal{V}_F$ with $\dim(U) = i$ and $\dim(W) = j$ (which exist if $k_i, k_j \geq 1, i \neq j$). As partition elements, it holds that $U \cap W = \{0\}$. The dimension formula then gives

$$\dim(U + W) = \dim(U) + \dim(W) = i + j \leq n.$$

Statement (iii) follows immediately. For (ii), if $k_i \geq 2$, we can choose distinct U, W of dimension i, implying $2i \leq n$, or $i \leq n/2$. Thus, if $i > n/2$, we must have $k_i \leq 1$. ■

Theorem 5.4.2. *Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a crooked quadratic function (where n is even if $p = 2$). Suppose that F has linearity $\mathcal{L}(F) = p^{\frac{n+\ell}{2}}$ with $\ell > n/2$. Then F has exactly $p - 1$ component functions with*

amplitude $p^{\frac{n+\ell}{2}}$, and all other non-trivial component functions have amplitude at most $p^{\frac{2n-\ell}{2}}$. In particular, any such function F has at most $p-1$ component functions with amplitude larger than $p^{3n/4}$.

Proof. The linearity $\mathcal{L}(F) = p^{\frac{n+\ell}{2}}$ implies the existence of a subspace $U \in \mathcal{V}_F$ with $\dim(U) = \ell$. Since $\ell > n/2$, the dimension bounds from the previous theorem imply that U is the unique subspace of this dimension ($k_\ell = 1$). This subspace corresponds to exactly $p-1$ component functions (see Remark 5.3.12), which are the unique components of maximal amplitude.

Now consider any other component function F_b associated with a subspace $W \in \mathcal{V}_F$. If $W \neq U$, the dimension bounds require that $\dim(U) + \dim(W) \leq n$. Substituting $\dim(U) = \ell$, we obtain $\dim(W) \leq n - \ell$. Therefore, the amplitude of F_b is bounded by

$$p^{\frac{n+(n-\ell)}{2}} = p^{\frac{2n-\ell}{2}}.$$

Finally, an amplitude strictly larger than $p^{3n/4}$ corresponds to a subspace dimension d satisfying $\frac{n+d}{2} > \frac{3n}{4}$, or $d > n/2$. As established, U is the only subspace with dimension exceeding $n/2$, so at most $p-1$ components can satisfy this condition. ∎

More conditions regarding the vector space partition, specifically using the so-called tail, i.e. the smallest dimension of a subspace in the partition, are provided in [7]. In the respective source [39], these results are proven for general p, so they also hold for odd characteristic and are translatable using Theorem 5.3.11.

5.4.2 Conditions on the number of non-bent components

In the following, the geometric connections of N_F with vector space partitions, blocking sets (in even characteristic), and hypersurfaces are leveraged to establish lower and upper bounds on the number of non-bent components of (crooked) quadratic functions, thereby also bounding the number of bent components.

From a cryptographic perspective, quantifying the number of bent components is of great interest. Particularly in characteristic two, bent components achieve optimal non-linearity and thus provide maximal resistance against linear cryptanalysis. Because perfectly bent vectorial Boolean functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2^n$ cannot exist [73, Corollary to Theorem 3.2], a fundamental cryptographic objective is to maximise the number of bent components (minimise the cardinality of N_F) to ensure security against linear attacks. By constraining the size of N_F, we outline its theoretical limits.

Conditions on the number of non-bent components via vector space partitions

Recalling the definition of N_F as the set of non-trivial, non-bent component functions and the fact that we discard zero-dimensional spaces in the vector space partition $\mathcal{V}_F$, together with Remark 5.3.12, we obtain the following result, generalising [7, Corollary 2.6].

Proposition 5.4.3. *Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a crooked quadratic function, and let $\mathcal{V}_F$ be its associated vector space partition. Then $|N_F| = (p-1)|\mathcal{V}_F|$.*

Due to the preceding proposition, establishing combinatorial bounds on the size of a vector space partition yields immediate bounds on $|N_F|$. To contextualise these bounds, we first define a highly uniform class of vector space partitions.

Definition 5.4.4 (Spread). A vector space partition $\mathcal{V}$ of $\mathbb{F}_p^n$ is called a *t-spread* if every subspace in $\mathcal{V}$ has the exact same dimension t.

It is well-known that a t-spread of $\mathbb{F}_p^n$ exists if and only if t divides n [92]. Furthermore, a counting argument demonstrates that a t-spread contains $\frac{p^n-1}{p^t-1}$ subspaces.

In the context of cryptography, spreads correspond to highly uniform Walsh spectra. For instance, a (crooked) quadratic APN function on $\mathbb{F}_2^n$ (n even) has a *classical Walsh spectrum* if all of its non-bent components have amplitude $2^{\frac{n+2}{2}}$. By Theorem 5.2.6, the vector space partition associated with such a function consists only of 2-dimensional subspaces, meaning it is a 2-spread. Currently, all known infinite families of quadratic APN functions $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$, for n even, exhibit this classical spectrum. Determining whether there exist infinite families with a non-classical Walsh spectrum remains a major open problem.

Bénéteau et al. [7] first introduced the technique of bounding the number of non-bent components $|N_F|$ by applying extremal bounds on the size of vector space partitions. In the following proposition, the first lower bounds based on [40, Theorem 1] are presented. However, when $k \mid n$, the minimum size of a partition with maximum dimension k is achieved if and only if the partition is a k-spread. If we are searching for functions with a non-classical Walsh spectrum, we are explicitly looking for partitions that are not k-spreads. This chapter tightens the bound for the non-classical case as compared to [7] by incorporating recent results by Năstase and Sissokho [71, Theorem 9]. They determined the second minimum size of such vector space partitions. Using Theorem 5.2.6 for $p = 2$ and Theorem 5.3.11 for $p > 2$, this leads to the following bounds.

Proposition 5.4.5. *Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a crooked quadratic function (where n is even if $p = 2$) with linearity $p^{\frac{n+k}{2}}$, where $k \leq n/2$. Set $n = sk + r$ with integers s, r and $0 \leq r < k$.*
If $k \mid n$, then

$$|N_F| \geq (p-1)\frac{p^n - 1}{p^k - 1}.$$

Furthermore, if the associated vector space partition $\mathcal{V}_F$ is not a k-spread, this bound strictly increases to

$$|N_F| \geq (p-1)\left(\frac{p^n - 1}{p^k - 1} + p^{\lceil \frac{k}{2} \rceil}\right).$$

If $k \nmid n$, then

$$|N_F| \geq (p-1)\left(1 + p^{\lceil \frac{k+r}{2} \rceil} + p^{k+r}\sum_{i=0}^{s-2} p^{ik}\right).$$

Similarly, we can adapt the maximum possible size of a vector space partition from [40, Theorem 1] to obtain an upper bound on $|N_F|$.

Proposition 5.4.6. *Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a crooked quadratic function (where n is even if $p = 2$) for which all components have amplitude at least $p^{\frac{n+k}{2}}$. Set $n = sk + r$ with integers s, r and $0 \leq r < k$. Then*

$$|N_F| \leq (p-1)\left(1 + p^{k+r}\sum_{i=0}^{s-2} p^{ik}\right).$$

The following is a consequence of [53, Proposition 7].

Proposition 5.4.7. *Let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a crooked quadratic function (where n is even if $p = 2$) with linearity $p^{\frac{n+k}{2}}$. Assume that F has another non-bent component with amplitude $p^{\frac{n+\ell}{2}}$ with $1 \leq \ell < n - k$ and $\ell \neq k$. Then*

$$|N_F| \geq (p-1)\left(p^k + p^\ell + 1\right).$$

These combinatorial limits on vector space partitions serve as a first tool to identify valid amplitude distributions. However, to extract even stronger bounds on the number of non-bent components, we must analyse the geometric structure of N_F.

Conditions on the number of non-bent components via blocking sets

We now turn our attention to the geometric structure of N_F. In the binary case with n even, for a quadratic function $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$, not necessarily crooked, we established that N_F forms a blocking set in $\mathrm{PG}(m-1,2)$ with respect to $(n/2)$-spaces (see Proposition 5.2.9). Moreover, it intersects every such subspace in an odd number of points.

It was observed in [84, Theorem 3] that the number of bent components of any vectorial Boolean function $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ is bounded from above by $2^m - 2^{m-n/2}$. Functions reaching this bound are called *maximum number of bent components* (MNBC) functions. This extremal case corresponds to the situation where the associated set N_F is a trivial blocking set with respect to $(n/2)$-spaces in $\mathrm{PG}(m-1,2)$, meaning it contains a projective subspace of dimension $m - n/2 - 1$.

Therefore, for any quadratic function F, not MNBC, the set N_F must be a non-trivial blocking set. To establish a lower bound on its size, we rely on the parity constraint: if such an odd-intersecting blocking set is relatively small, it must necessarily be minimal. We formalise this with the following two results, which classify the sizes and structures of such minimal sets.

Proposition 5.4.8 ([7, Proposition 3.6]). *Let B be a blocking set of $\mathrm{PG}(n,2)$ with respect to k-spaces, such that B intersects every k-space in an odd number of points. If $|B| \leq 2^{m-k+2} - 2$ then B is a minimal blocking set.*

Proposition 5.4.9 ([37, Theorem 1.4(iii)]). *In $\mathrm{PG}(n,2)$, $n \geq 3$, the smallest non-trivial blocking sets with respect to k-spaces, $1 \leq k \leq n-2$, have size $2^{n-k+1} + 2^{n-k-1} + 2^{n-k-2} - 1$ and are cones with vertex an $(n-k-3)$-space π_{n-k-3} and base the set of points on the edges of a tetrahedron in a solid skew to π_{n-k-3}.*

From this, Bénéteau et al. [7, Theorem 3.7] derived the following lower bound on the number of non-bent components N_F of a quadratic function F, not MNBC, under the assumptions n even and $n \leq 2m-6$:

$$|N_F| \geq 2^{m-n/2} + 2^{m-n/2-2} + 2^{m-n/2-3} - 1.$$

Subsequently, in [7, Open Problem 6.4], the question was posed whether there exists a quadratic function F, not MNBC, reaching this exact bound, i.e. whether the bound is sharp. The answer to this open problem is negative. We demonstrate this using the geometric structure of the cone described in Proposition 5.4.9. First, recall that the span $\langle S \rangle$ of a subset S in $\mathrm{PG}(n,p)$ represents the smallest subspace in $\mathrm{PG}(n,p)$ containing all elements of S.

Theorem 5.4.10. *Let n be even with $n \leq 2m-6$. There does* not *exist a non-MNBC quadratic function $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ with $|N_F| = 2^{m-n/2} + 2^{m-n/2-2} + 2^{m-n/2-3} - 1$.*

Proof. Assume, for the sake of contradiction, that there exists such a function F for which $|N_F| = 2^{m-n/2} + 2^{m-n/2-2} + 2^{m-n/2-3} - 1$. Since F is not an MNBC function, it follows that N_F must be a non-trivial blocking set. Then, by Proposition 5.4.8 and Proposition 5.4.9, N_F is a cone in $\mathrm{PG}(m-1,2)$ with $(m-n/2-4)$-dimensional vertex $\pi_{m-n/2-4}$ and base B_0 consisting of the ten points on the edges of a tetrahedron in a solid $\mathcal{S} \cong \mathrm{PG}(3,2)$ such that $\mathcal{S} \cap \pi_{m-n/2-4} = \emptyset$; see Figure 5.1.

Thus, $N_F = \bigcup_{P \in B_0} \langle \pi_{m-n/2-4}, P \rangle$ and define $\mathcal{K} = \langle N_F \rangle$. It then holds that

$$\dim(\mathcal{K}) = \dim(\pi_{m-n/2-4}) + \dim(\mathcal{S}) - \dim(\emptyset) = \left(m - \frac{n}{2} - 4\right) + 3 - (-1) = m - \frac{n}{2}.$$

In $\mathcal{S}$, let P_1, P_2, P_3, P_4 be four non-coplanar points, identified without loss of generality as the vertices of our tetrahedron. Then B_0 contains these points and the points on the edges between them (their pairwise sums). Define the line $\mathcal{L} = \langle P_1, P_2 + P_3 \rangle = \{P_1, P_2 + P_3, P_1 + P_2 + P_3\}$. We have $P_1, P_2 + P_3 \in B_0$, but $P_1 + P_2 + P_3 \notin B_0$ as it lies on a face of the tetrahedron. Because $\mathcal{L} \subseteq \mathcal{S}$, it is skew to the vertex $\pi_{m-n/2-4}$ and hence it can only intersect N_F at the base B_0. Thus, $|\mathcal{L} \cap N_F| = 2$.

Now, as we work in $\mathrm{PG}(m-1,2)$ and $\dim(\mathcal{K}) = m - \frac{n}{2}$, there exists a subspace $\pi_{n/2-2}$ of dimension $\frac{n}{2}-2$ such that $\mathcal{K} \cap \pi_{n/2-2} = \emptyset$. Indeed, the dimensions add up:

$$\dim(\langle \mathcal{K}, \pi_{n/2-2} \rangle) = \left(m - \frac{n}{2}\right) + \left(\frac{n}{2} - 2\right) + 1 = m - 1 = \dim(\mathrm{PG}(m-1,2)).$$

We can then define $\mathcal{U} = \langle \mathcal{L}, \pi_{n/2-2} \rangle$ and observe that $\dim(\mathcal{U}) = 1 + \left(\frac{n}{2} - 2\right) + 1 = \frac{n}{2}$. By definition, any point $X \in \mathcal{U}$ lies on a line connecting a point in $\mathcal{L}$ and a point in $\pi_{n/2-2}$. Because $\pi_{n/2-2}$ is skew to $\mathcal{K}$ and $\mathcal{L} \subseteq \mathcal{K}$, the only points in $\mathcal{U}$ that also belong to $\mathcal{K}$ are the points of $\mathcal{L}$ itself. Therefore, $\mathcal{U} \cap \mathcal{K} = \mathcal{L}$. Hence,

$$|\mathcal{U} \cap N_F| = |\mathcal{U} \cap (\mathcal{K} \cap N_F)| = |\mathcal{L} \cap N_F| = 2.$$

So, we have constructed an $(n/2)$-space $\mathcal{U}$ that intersects N_F in an even number of points, which contradicts Proposition 5.2.9. This completes the proof. ∎

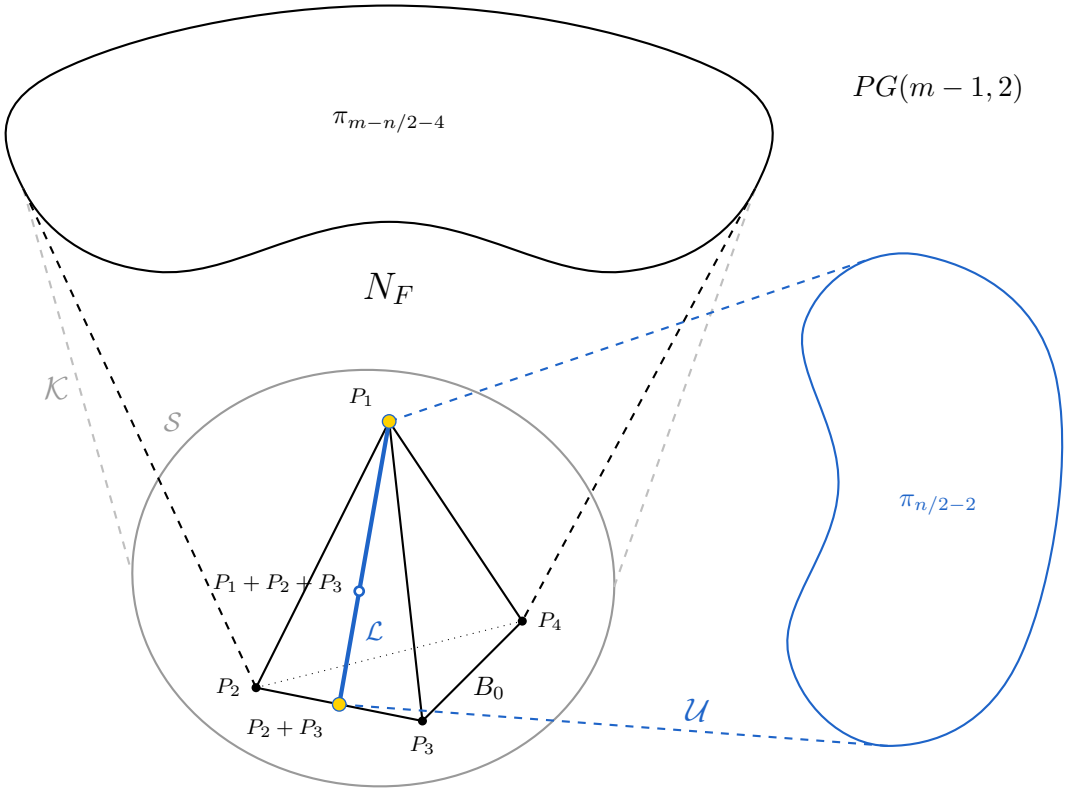


Figure 5.1: Geometric configuration constructed in the proof of Theorem 5.4.10.

Combining this finding with the known lower bound naturally leads to a strict inequality for the size of N_F.

Corollary 5.4.11. *Let n be even with $n \leq 2m-6$ and $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a quadratic function. If F is not an MNBC function, then*

$$|N_F| > 2^{m-n/2} + 2^{m-n/2-2} + 2^{m-n/2-3} - 1.$$

Proof. By [7, Theorem 3.7], we know that $|N_F| \geq 2^{m-n/2} + 2^{m-n/2-2} + 2^{m-n/2-3} - 1$. However, the preceding theorem demonstrates that equality cannot hold. Thus, the strict bound follows. ∎

For the case $n = 2$, we show that if a quadratic function $F\colon \mathbb{F}_2^2 \to \mathbb{F}_2^m$ is not MNBC, its associated set of non-bent components N_F necessarily comprises the entire projective space $\mathrm{PG}(m-1,2)$.

Proposition 5.4.12. *Let $F\colon \mathbb{F}_2^2 \to \mathbb{F}_2^m$ be a quadratic function. If F is not an MNBC function, then $N_F = \mathrm{PG}(m-1,2)$.*

Proof. Assume, for the sake of contradiction, that $N_F \neq \mathrm{PG}(m-1,2)$. Then there exists a point $P \in \mathrm{PG}(m-1,2)$ such that $P \notin N_F$.

By Proposition 5.2.9, N_F is a blocking set that intersects every line in $\mathrm{PG}(m-1,2)$ in an odd number of points. Consider the set of all lines passing through the point P. In $\mathrm{PG}(m-1,2)$, there are exactly $2^{m-1}-1$ such lines, and they partition the set of points $\mathrm{PG}(m-1,2) \setminus \{P\}$.

Let $\mathcal{L}$ be any line through P. Since a line in $\mathrm{PG}(m-1,2)$ contains exactly 3 points and $P \notin N_F$, the intersection $\mathcal{L} \cap N_F$ can contain at most 2 points. However, due to the odd intersection property, we must have $|\mathcal{L} \cap N_F| = 1$. Thus, every line through P contains exactly one point of N_F. Because the lines through P partition the remaining space and each contains exactly one point of N_F, the total number of points in N_F equals the number of lines through P. That is, $|N_F| = 2^{m-1}-1$.

Now, let Q_1 and Q_2 be two distinct points in N_F, and let $\langle Q_1, Q_2\rangle$ be the line spanned by them. Since $\langle Q_1, Q_2\rangle$ must intersect N_F in an odd number of points and already contains both Q_1 and Q_2, it must also contain the third point on the line. Therefore, $|\langle Q_1, Q_2\rangle \cap N_F| = 3$, which implies $\langle Q_1, Q_2\rangle \subseteq N_F$.

Because any line connecting two points in N_F is entirely contained within N_F, the set N_F forms a projective subspace. A subspace of size $2^{m-1}-1$ in $\mathrm{PG}(m-1,2)$ is a hyperplane, meaning it has dimension $m-2$. By definition, a blocking set in $\mathrm{PG}(m-1,2)$ with respect to 1-spaces that contains an $(m-2)$-space is a trivial blocking set. However, as established earlier, if F is not an MNBC function, N_F must be a non-trivial blocking set. This is a contradiction. ∎

While viewing N_F as an odd-intersecting blocking set has proven useful for establishing first lower bounds on the number of non-bent components, we were able to derive a lower bound that is not only stronger but also sharp by viewing N_F as a projective hypersurface.

Conditions on the number of non-bent components via hypersurfaces

The strict inequality established in Corollary 5.4.11 naturally raises the problem of determining a sharp lower bound. Given the highly constrained algebraic nature of N_F, its possible cardinalities are expected to be sparsely distributed. This suggests that the true minimum size of N_F for a non-MNBC function may be significantly larger. This motivated the search for another geometric structure which enables a stronger lower bound. By characterising N_F as a projective hypersurface, we are able to establish a substantially stronger – and importantly, sharp – lower bound under slightly weaker assumptions.

We will need the notion of Reed-Muller codes.

Definition 5.4.13 (Reed-Muller code). The r-th order *Reed-Muller code* of length 2^m, denoted by $\mathrm{RM}(r,m)$, is the linear code consisting of the evaluation vectors, i.e. the vectors of length 2^m recording the outputs of the function for all possible inputs, of all Boolean functions $f\colon \mathbb{F}_2^m \to \mathbb{F}_2$ of algebraic degree at most r.

Such codes are extensively studied in coding theory, and the possible Hamming weights of codewords in a Reed-Muller code are highly restricted. The weights lying between the minimum distance and twice the minimum distance were completely determined in [46].

Lemma 5.4.14 ([46, Theorem 1]). *Let $\mathrm{RM}(r,m)$ be the Reed-Muller code of order r and length 2^m, and let $d = 2^{m-r}$ denote its minimum distance. The only possible Hamming weights w of the codewords in $\mathrm{RM}(r,m)$ that lie in the range $d \leq w < 2d$ are given by*

$$w = 2^{m-r+1} - 2^{m-r+1-i},$$

for some integer $1 \leq i \leq \max\left(\min(r, m-r), \lfloor \frac{m-r+2}{2} \rfloor\right)$.

By associating the homogeneous polynomial defining our hypersurface with a codeword in the code $\mathrm{RM}(n/2,m)$, we obtain the following lower bound in characteristic two.

Theorem 5.4.15. *Let n be even with $n \leq 2m-4$ and $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a quadratic function. If F is not an MNBC function, then*

$$|N_F| \geq 2^{m-n/2} + 2^{m-n/2-1} - 1.$$

Proof. By Theorem 5.2.12, the set of non-bent components N_F is the zero set of a homogeneous polynomial $P(b_1, \ldots, b_m)$ of degree $n/2$ with coefficients in $\mathbb{F}_2$. That is, for any non-zero $b \in \mathbb{F}_2^m$, the component F_b is non-bent if and only if $P(b) = 0$. Equivalently, the component F_b is bent precisely when $P(b) = 1$. Since P is a homogeneous polynomial of degree $n/2 \geq 1$, it evaluates to 0 at the origin, meaning $P(0) = 0$. This corresponds to the trivial component F_0, a component that is never bent.

We can view P as a Boolean function $P\colon \mathbb{F}_2^m \to \mathbb{F}_2$ because it evaluates an m-dimensional binary vector b to a single binary value in $\mathbb{F}_2$. Since its algebraic degree is $n/2$, its evaluation vector forms a codeword in the Reed-Muller code $\mathrm{RM}(n/2, m)$. Let us define the Boolean function $G\colon \mathbb{F}_2^m \to \mathbb{F}_2$ by $G(b) = 1 + P(b)$. Because G differs from P only by a constant term, it also has an algebraic degree of $n/2$, meaning $G \in \mathrm{RM}(n/2, m)$.

The Hamming weight of G, denoted $\mathrm{wt}_H(G)$, is the number of inputs $b \in \mathbb{F}_2^m$ for which $G(b) = 1$. By definition, $G(b) = 1$ if and only if $P(b) = 0$. Thus, G evaluates to 1 exactly on the elements of N_F and at the origin $b = 0$. Consequently, the weight of the codeword G is precisely

$$\mathrm{wt}_H(G) = |N_F| + 1.$$

By assumption, F is not an MNBC function. Therefore, the weight of the codeword G must be strictly larger than the minimum weight $2^{m-n/2}$ of $\mathrm{RM}(n/2, m)$. By Lemma 5.4.14, since $n \leq 2m-4$ allows for $i = 2$, the next possible weight for a codeword in $\mathrm{RM}(n/2, m)$ is $2^{m-n/2} + 2^{m-n/2-1}$ (which is equivalent to $2^{m-n/2+1} - 2^{m-n/2-1}$). Therefore, we must have

$$\mathrm{wt}_H(G) \geq 2^{m-n/2} + 2^{m-n/2-1}.$$

Substituting $\mathrm{wt}_H(G) = |N_F| + 1$ into this inequality and rearranging yields

$$|N_F| \geq 2^{m-n/2} + 2^{m-n/2-1} - 1.$$

This concludes the proof. ∎

Furthermore, the new lower bound is sharp, as is shown in the next example.

Example 5.4.16. Consider the Gold monomial $F\colon \mathbb{F}_2^4 \to \mathbb{F}_2^4$ defined by $F(x) = x^3$, where we identify the vector space with the finite field $\mathbb{F}_{16}$.

As noted in [7, Example 3.10], for the crooked function $F(x) = x^3$ (see Example 5.3.4), the set of non-bent components N_F corresponds to the set of non-zero cubes in $\mathbb{F}_{16}$. Furthermore, since $n/2 = 2$ is even, the same example establishes that N_F does not contain an $(n/2)$-dimensional subspace, meaning F is not an MNBC function.

In the multiplicative group $\mathbb{F}_{16}^*$, which contains exactly 15 elements, the image of the mapping $x \mapsto x^3$ is precisely the set of non-zero cubes. The size of this image is given by

$$\frac{15}{\gcd(15, 3)} = \frac{15}{3} = 5.$$

Therefore, there are 5 non-zero cubes, which implies that $|N_F| = 5$.

Evaluating the lower bound from Theorem 5.4.15 for $n = m = 4$ yields:

$$|N_F| \geq 2^{4-2} + 2^{4-3} - 1 = 4 + 2 - 1 = 5.$$

Since $|N_F|$ is exactly 5, the Gold monomial on $\mathbb{F}_2^4$ achieves the bound perfectly, demonstrating its sharpness. □

While the lower bound derived via the hypersurface characterisation is sharp, explicitly constructing functions that achieve this exact bound for $m = n > 4$ remains a challenging open problem. Monomial functions cannot achieve this bound for larger n because their number of non-bent components (which equals $2^{n/2}+2^{n/2-1}-1$ in the minimal case) must divide 2^n-1. At the same time, manually constructing a custom quadratic function requires solving a system of algebraic equations whose complexity scales rapidly as the dimension increases. Nonetheless, as demonstrated by the case $n = m = 4$, this lower bound is the most general sharp bound applicable to the entire parameter range $n \leq 2m-4$. Determining whether quadratic functions can actually reach this minimum for specific higher dimensions, or whether stricter bounds emerge in those cases, provides an interesting direction for future research.

The previous best bound established in [7, Theorem 3.7] relied on the classification of minimal blocking sets [37]. As such, it requires the dimension parameters to satisfy $n \leq 2m - 6$ (alongside n being even). In the most cryptographically relevant case where the domain and codomain are of equal size ($m = n$), this restricts the applicability of the bound to $n \geq 6$. In contrast, our new bound relies on the weight distribution of Reed-Muller codes. To guarantee the existence of the second weight via Lemma 5.4.14, this approach only requires $n \leq 2m - 4$. For $m = n$, this condition simplifies to $n \geq 4$, successfully lowering the dimensional requirement while simultaneously yielding a strictly larger bound. Also note that the case $n = 2$ is covered by Proposition 5.4.12.

Furthermore, we can compare the magnitude of the bounds algebraically. Let $d = 2^{m-n/2}$ denote the theoretical minimum size of a blocking set with respect to $(n/2)$-spaces (which is the minimum non-zero weight of the associated Reed-Muller code). The previous bound can be rewritten in terms of d as

$$|N_F| \geq d + \frac{d}{4} + \frac{d}{8} - 1 = \frac{11}{8}d - 1.$$

Our new bound is given by

$$|N_F| \geq d + \frac{d}{2} - 1 = \frac{3}{2}d - 1.$$

Because $\frac{3}{2} = \frac{12}{8} > \frac{11}{8}$, the new bound is strictly larger and guarantees a higher minimum number of non-bent components for any quadratic function that is not MNBC.

It is known that (crooked) quadratic APN functions cannot be MNBC functions in characteristic two [69, Theorem 2]. Applying the theorem above to the case $n = m$ yields an (improved over [7, Theorem 3.9]) general upper bound on the number of bent components for quadratic APN functions.

Corollary 5.4.17. *Let $n \geq 4$ be even and $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a quadratic APN function. Then*

$$|N_F| \geq 2^{n/2} + 2^{n/2-1} - 1.$$

In other words, the number of bent component functions of F is less than or equal to $2^n - 2^{n/2} + 2^{n/2-1}$.

In the odd characteristic setting ($p > 2$), we find the same algebraic structure that we used in the binary case: the set N_F forms a projective hypersurface. However, its defining homogeneous polynomial now has degree n, rather than $n/2$. This higher degree prevents us from establishing a general lower bound on the size of N_F using the weight distribution of Reed-Muller codes in the cryptographically significant case where $m = n$. Nevertheless, this hypersurface characterisation allows us to deduce a geometric property regarding the intersection of N_F with lines in the dual space.

Theorem 5.4.18. *Let $p > n \geq 2$ and let $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be a quadratic function. Consider a line $\mathcal{L} \subseteq \mathrm{PG}(n-1,p)^*$. Let $k = |\mathcal{L} \cap N_F|$ be the number of non-bent components on this line. Then k must satisfy one of the following:*

1. *$k \leq n$,*
2. *$k = p + 1$ (the entire line is contained in N_F).*

Proof. Let $P(b_1, \ldots, b_n)$ be the homogeneous polynomial of degree n defining N_F, as established in Lemma 5.3.15. Let the line $\mathcal{L}$ be spanned by two distinct points $\langle u \rangle$ and $\langle v \rangle$ in $\mathrm{PG}(n-1,p)^*$. The points on $\mathcal{L}$ can be parametrised by pairs $(s,t) \in \mathbb{F}_p^2 \setminus \{(0,0)\}$ (up to scalar equivalence) as $\langle su + tv \rangle$.

We evaluate P on the line $\mathcal{L}$ by defining the polynomial in two variables:

$$R(s,t) = P(su + tv).$$

Since P is homogeneous of degree n, $R(s,t)$ is a homogeneous polynomial of degree n in variables s and t. The intersection points in $\mathcal{L} \cap N_F$ correspond exactly to the roots of $R(s,t) = 0$ on the projective line $\mathcal{L} \cong \mathrm{PG}(1,p)$.

If $R(s,t)$ is the zero polynomial, P vanishes for every point on the line $\mathcal{L}$. Therefore, $\mathcal{L} \subseteq N_F$, and the number of intersection points is the total number of points on a projective line over $\mathbb{F}_p$, which is $k = p + 1$. On the other hand, if $R(s,t)$ is not identically zero, then it follows from the fundamental theorem of algebra (adapted for homogeneous polynomials) that a non-zero homogeneous polynomial of degree n over a field has at most n zeroes in the projective space $\mathrm{PG}(1,p) \cong \mathcal{L}$. Therefore, $k \leq n$.

Since we assumed $p > n$, these two cases are disjoint, proving the claim. ∎

Note that if $p \leq n$, the second case ($k = p + 1$) is effectively absorbed into the first bound $k \leq n$, which justifies the condition $p > n$ for this theorem.

While this intersection constraint applies to any quadratic function, it becomes a particularly useful tool for ruling out hypothetical amplitude distributions when we restrict our focus back to crooked quadratic functions.

Example 5.4.19. Let $m = n = 4$ and $p = 5$. The projective space $\mathrm{PG}(3,5)$ contains $(5^4 - 1)/(5 - 1) = 156$ points.

We construct a vector space partition $\mathcal{V}$ of $\mathbb{F}_5^4$ of type $[1^{125}, 3^1]$. This partition consists of one hyperplane U and 125 vector lines. Such a partition exists because the complement of the hyperplane, $\mathbb{F}_5^4 \setminus U$, contains $5^4 - 5^3 = 500$ non-zero vectors, which can be partitioned into $500/(5-1) = 125$ disjoint vector lines. It is easily verified that the conditions of Theorem 5.4.1 and Theorem 5.4.2 are satisfied.

Now, consider the size of the set of non-bent components N_F in the dual projective space $\mathrm{PG}(3,5)^*$. Since there is a one-to-one correspondence between the subspaces in $\mathcal{V}$ and the *projective* points in N_F (see Equation (5.5), Proposition 5.3.7 and Theorem 5.3.10 because $m = n$), we have

$$|N_F|_{\text{proj}} = |\mathcal{V}| = 1 + 125 = 126.$$

So, the complement set $N_F^c = \mathrm{PG}(3,5)^* \setminus N_F$ has size $156 - 126 = 30$.

Let Q be an arbitrary point in this complement N_F^c. The number of lines passing through a fixed point in $\mathrm{PG}(3,5)^*$ is given by the Gaussian coefficient corresponding to the number of 2-dimensional subspaces containing a fixed 1-dimensional subspace in $\mathbb{F}_5^4$. This is equivalent to counting the number of 1-dimensional subspaces in the quotient space of dimension $4 - 1 = 3$:

$$\begin{bmatrix} 4-1 \\ 2-1 \end{bmatrix}_5 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}_5 = \frac{5^3 - 1}{5 - 1} = 31.$$

We now check if it is possible for every line through Q to contain another point from the complement. Excluding Q itself, there are exactly $30 - 1 = 29$ remaining points in N_F^c. Even in the worst-case scenario where each of these 29 points lies on a different line passing through Q, they can determine at most 29 lines.

Since there are 31 lines in total passing through Q, there must be at least $31 - 29 = 2$ lines passing through Q that contain no other points from the complement. Let $\mathcal{L}$ be one such line. As $\mathcal{L}$ contains Q (which is in N_F^c) and no other points from N_F^c, the intersection with the complement is exactly $|\mathcal{L} \cap N_F^c| = 1$.

A line in $\mathrm{PG}(3,5)^*$ contains $p+1=6$ points. Therefore, the intersection with N_F equals

$$|\mathcal{L} \cap N_F| = |\mathcal{L}| - |\mathcal{L} \cap N_F^c| = 6 - 1 = 5.$$

However, Theorem 5.4.18 states that for $p > n$, the intersection size k of a line with N_F must satisfy either $k \leq n = 4$ or $k = p + 1 = 6$. The value $k = 5$ falls into the forbidden gap. Hence, no crooked quadratic function can have an amplitude distribution corresponding to the partition type $[1^{125}, 3^1]$. □

This example highlights that the existence of a valid vector space partition is by no means sufficient for an amplitude distribution to be achievable by a crooked quadratic function. The example can be generalised, as is shown next.

Corollary 5.4.20. *Let $p > n \geq 3$. There exists no crooked quadratic function $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ with an amplitude distribution corresponding to the vector space partition type $[1^{p^d\theta_{n-d-1}}, d^1]$, where $\theta_k = \frac{p^{k+1}-1}{p-1}$, for any $2 \leq d \leq n-1$. In other words, a partition consisting of exactly one subspace of dimension $2 \leq d \leq n-1$, with the rest of the space partitioned into lines, cannot be realised by such a function.*

Proof. Assume, for the sake of contradiction, that such a function F exists. The associated vector space partition $\mathcal{V}_F$ consists of exactly one subspace U of dimension d and a set of lines $\{L_i\}$ partitioning the complement $\mathbb{F}_p^n \setminus U$. The number of lines is determined by the size of the complement:

$$|\{L_i\}| = \frac{p^n - p^d}{p-1} = p^d \frac{p^{n-d}-1}{p-1} = p^d\theta_{n-d-1}.$$

In the dual projective space $\mathrm{PG}(n-1,p)^*$, the set N_F contains exactly 1 point corresponding to U and $p^d\theta_{n-d-1}$ points corresponding to the lines. We compute the size of the complement set N_F^c. Using the identity $\theta_{n-1} = \theta_{d-1} + p^d\theta_{n-d-1}$, we obtain

$$|N_F^c| = \theta_{n-1} - |N_F| = (\theta_{d-1} + p^d\theta_{n-d-1}) - (1 + p^d\theta_{n-d-1}) = \theta_{d-1} - 1.$$

Since $d \geq 2$, it holds that $\theta_{d-1} \geq p + 1 > 1$, so the complement is non-empty. Let Q be an arbitrary point in N_F^c. The number of lines passing through Q in $\mathrm{PG}(n-1,p)^*$ equals θ_{n-2}.

We now determine if there exists a line through Q that intersects the complement N_F^c only at Q. The number of points in $N_F^c \setminus Q$ is equal to $R = |N_F^c| - 1 = \theta_{d-1} - 2$. By the pigeonhole principle, the number of lines passing through Q that contain *no* other points of N_F^c is at least

$$\theta_{n-2} - R = \theta_{n-2} - (\theta_{d-1} - 2).$$

Since $d \leq n-1$, it implies $d - 1 \leq n - 2$ and thus $\theta_{d-1} \leq \theta_{n-2}$. Therefore, the number of such lines is strictly positive:

$$\theta_{n-2} - \theta_{d-1} + 2 \geq 2.$$

Hence, there exist at least two lines through $Q \in N_F^c$ that are entirely contained in N_F except for the point Q. Let $\mathcal{L}$ be one such line. Its intersection with N_F has cardinality

$$|\mathcal{L} \cap N_F| = |\mathcal{L}| - |\mathcal{L} \cap N_F^c| = (p+1) - 1 = p.$$

However, according to Theorem 5.4.18, for $p > n$, the intersection size $k = |\mathcal{L} \cap N_F|$ must satisfy either $k \leq n$ or $k = p + 1$. Since $n < p < p + 1$, the value $k = p$ is strictly forbidden. This contradiction implies that no such function F can exist. ■

We conclude that the algebraic structure of (crooked) quadratic functions imposes additional geometric constraints on the set N_F which can rule out partition types that are otherwise combinatorially sound.

5.5 Summary of structural connections

In this chapter, we have established a geometric framework for analysing the Walsh spectrum of cryptographic functions. We demonstrated that for specific classes of functions, the spectral properties are not arbitrary but are strictly governed by the geometry of the underlying vector space and its dual projective space.

While the fundamental link to vector space partitions remains relatively consistent across characteristics, the geometric nature of the set of non-bent components N_F reveals distinct properties. In characteristic two, N_F acts as both a blocking set and a projective hypersurface, enabling us to establish a sharp lower bound on its size. In odd characteristic, the hypersurface characterisation yields stringent line intersection constraints. The structural correspondences are summarised in Table 5.1.

Table 5.1: Comparison of structural connections for the Walsh spectrum between characteristic two and odd characteristic.

	Characteristic two ($p = 2$)	**Odd characteristic ($p > 2$)**
Function	Crooked quadratic (APN if $m = n$)	Crooked quadratic
Partition structure ($m = n$)	Partition $\mathcal{V}_F$ composed of subspaces V_b	Partition $\mathcal{V}_F$ composed of subspaces $V_{\langle b \rangle}$
Amplitude	$2^{\frac{n+\dim(V_b)}{2}}$	$p^{\frac{n+\dim(V_{\langle b \rangle})}{2}}$
Geometry of N_F	Blocking set in $\mathrm{PG}(m-1,2)$ and projective hypersurface in $\mathrm{PG}(m-1,2)^*$	Projective hypersurface in $\mathrm{PG}(m-1,p)^*$
Derived constraint	Sharp lower bound: $\lvert N_F \rvert \geq 2^{m-n/2} + 2^{m-n/2-1} - 1$	Line intersection size k satisfies $k \leq n$ or $k = p+1$ (for $p > n$)

These structural constraints provide powerful necessary conditions for the existence of functions with specific Walsh spectra and can serve as a filter to narrow down the search space for functions with optimal cryptographic properties.

6 Conclusion and future work

> *Every wall is a door.*
> – Ralph Waldo Emerson

The central theme of this thesis has been the fruitful interplay between cryptography and finite geometry. Throughout this text, we have demonstrated that translating between algebra and finite geometry provides new insights, simplifies analyses, and yields solutions to cryptographic problems that are otherwise difficult to tackle in a strictly algebraic setting. This geometric lens has proven to be a useful tool in both asymmetric post-quantum cryptography and the construction of secure symmetric ciphers.

In the first part, we explored the foundations of code-based cryptography, motivated by the imminent threat of large-scale quantum computing. We detailed how the vulnerability of the classical GPT cryptosystem to structural attacks necessitates the search for new families of maximum-rank distance codes. After initially establishing some algebraic representations of rank-metric codes, we transitioned to their geometric construction via scattered subspaces. By formulating the exact geometric constraints required for these constructions to yield maximum-rank distance codes (and showing that they are inequivalent to existing codes under a certain condition), we demonstrated how finite geometry aids the discovery and classification of inequivalent code families. In addition, we included a result showcasing the reverse direction: successfully deducing the intersection numbers of linear sets from the rank distribution of the associated maximum-rank distance code.

In the second part, we shifted our focus to the security of some building blocks in symmetric cryptography, specifically examining the Walsh spectrum of quadratic vectorial functions. We successfully extended the geometric framework linking these functions to vector space partitions to fields of odd characteristic using the class of crooked quadratic functions. In addition, by characterising the underlying structure of the set of non-bent components as a projective hypersurface, we resolved an open problem in the binary case and established novel, sharp lower bounds on its cardinality. Furthermore, this characterisation yielded the first constraint of its kind in odd characteristic: by showing how this hypersurface intersects projective lines in the dual space, we obtained specific geometric configurations of the set of non-bent components.

6.1 Future work

While various active areas of research have already been mentioned throughout the text, we propose the following directions for future research that build upon the results presented in this thesis.

(i) Constructing new maximum h-scattered subspaces remains a notably difficult open problem in finite geometry. In addition, before the codes associated with these geometric constructions can be deployed in public-key cryptosystems, efficient polynomial-time decoding algorithms must be developed. Currently, such codes lack the algebraic structure required by existing polynomial-time decoders. Therefore, entirely novel decoding strategies must be devised to make the newly constructed codes practically relevant.

(ii) In the context of symmetric cryptography, a long-standing open problem is the discovery of an infinite family of binary quadratic APN functions that exhibits a non-classical Walsh spectrum. The exact lower bounds on the size of the non-bent component set N_F established in this work can be leveraged to limit the computational search space. In particular, by incorporating recent results by

Năstase and Sissokho [71] on the second minimum size of vector space partitions, Proposition 5.4.5 strictly improves the lower bound presented in [7] for this goal. These tightened constraints act as a filter, substantially optimising future computational searches for these functions.

(iii) A promising theoretical extension leverages the structural classification of low-weight Reed-Muller codewords to further analyse the geometry of the set of non-bent components N_F. In Chapter 5, applying the Kasami-Tokura bound [46] for RM$(n/2, m)$ to the indicator polynomial $G = 1 + P$, where the zero set of P defines N_F and $\mathrm{wt}(G) = |N_F \cup \{0\}|$, established the exact minimal cardinality of N_F. Extending this methodology, the results in [47] classify the weights and algebraic structures of Reed-Muller codewords up to $2.5d$, where $d = 2^{m-n/2}$ is the minimum distance. Applying this classification to G yields discrete cardinality gaps and structural restrictions below $2.5d - 1$. Furthermore, if we restrict $|N_F| < 2d - 1$, the original classification [46] gives explicit factorisations of G, which take the general form

$$G = (1 + L_1) \cdots (1 + L_t)Q,$$

where the L_i are independent affine linear polynomials and Q is a residual polynomial. Here, t depends on the corresponding factorisation; in particular, it holds that $t = n/2 - \mu$ for some $\mu \geq 2$.

When a factorisation takes this form, evaluations yielding $G(b) = 1$ are constrained by the simultaneous behaviour of the affine factors $1 + L_i$, and hence by the corresponding affine hyperplane arrangement. For the special codewords arising here, however, $G = 1 + P$ with P a homogeneous Pfaffian polynomial. Therefore, the classification of low-weight Reed-Muller codewords cannot be applied naively: the additional homogeneity constraint on $G + 1$ may rule out some of the factorisations. Understanding how the Pfaffian condition restricts the possible low-weight factorisations of G, and what geometric consequences this imposes on N_F, remains an interesting direction for future research.

(iv) Another direction for future research is generalising the framework developed in this thesis along two natural axes: moving to vectorial functions of algebraic degree $r > 2$, and extending the bounding methodology for the number of non-bent components to odd characteristics. The former is of significant cryptographic importance, as higher-degree functions provide improved resistance against so-called algebraic attacks. Extending this theory can be approached from both an algebraic and a geometric perspective. Algebraically, we can try to use r-th order Reed-Muller codes RM(r, n) and their p-ary generalisations. However, note that determining the exact weight distributions of higher-order codes remains a difficult open problem. The geometric perspective, therefore, offers an interesting alternative. Extending the study of vector space partitions and projective hypersurfaces to higher-degree varieties could yield independent structural bounds on the non-bent components, potentially bypassing the need for exact Reed-Muller weight enumerators. While establishing universal constraints requires further breakthroughs in classifying these algebraic varieties, deriving partial results remains a concrete and viable research direction.

Part III

Appendix

A Nederlandstalige samenvatting

Cryptografie vormt de hoeksteen van onze moderne digitale infrastructuur. Een eerste onderscheid kan gemaakt worden tussen symmetrische en asymmetrische cryptografie, waarbij verzender en ontvanger in het eerste geval over dezelfde sleutel beschikken en in het tweede geval de publieke sleutel voor encryptie verschilt van de geheime sleutel voor decryptie. Door de vooruitgang die geboekt wordt op het vlak van kwantumcomputers, die de meest gebruikte asymmetrische cryptosystemen vandaag de dag kunnen kraken, is er dringend nood aan post-kwantum alternatieven. Tegelijkertijd blijven ook symmetrische systemen onderhevig aan steeds geavanceerdere cryptanalyse. Deze masterproef onderzoekt belangrijke vraagstukken binnen beide domeinen, waarbij een vruchtbare wisselwerking met eindige meetkunde centraal staat.

Deel I: Code-gebaseerde cryptografie

Het eerste deel van deze thesis focust op asymmetrische post-kwantum cryptosystemen gebaseerd op foutverbeterende codes. Door het doelbewust introduceren van fouten in een boodschap die alleen de legitieme ontvanger kan verbeteren, is veilige data-overdracht mogelijk. Het klassieke McEliece-raamwerk, dat gebruikmaakt van de Hamming-metriek, kampt in de praktijk met onwerkbaar grote publieke sleutels. Om dit probleem te verhelpen, introduceerden Gabidulin, Paramonov en Tretjakov in 1991 het GPT-cryptosysteem [29] dat overstapt naar de rangmetriek (Eng.: *rank metric*). Door gebruik te maken van *maximum-rank distance* (MRD) codes (die een optimale foutverbeteringscapaciteit hebben voor gegeven codelengte en -dimensie), en in het bijzonder de efficiënt decodeerbare Gabidulin-codes, konden de sleutelgroottes drastisch verkleind worden.

De zwakte van Gabidulin-codes

De veiligheid van het GPT-cryptosysteem steunt op de aanname dat de publieke sleutel, een door middel van een vervormingsmatrix (Eng.: *scrambler matrix*) S en een verstoringsmatrix (Eng.: *distortion matrix*) X gemaskeerde versie van de geheime generatormatrix van een Gabidulin code, ononderscheidbaar is van de generatormatrix van een willekeurige code. Het decodeerprobleem voor een willekeurige code, dat is het kunnen vinden van het foutloze bericht uit een ontvangen bericht met fouten, is immers met grote kans NP-compleet [32]. Gabidulin-codes bezitten echter een zeer sterke algebraïsche structuur: ze zijn $\mathbb{F}_{q^m}$-lineair en hun generatormatrix wordt opgebouwd door opeenvolgende Frobenius-machten van een generatorvector te nemen. Deze interne structuur bleek de achilleshiel van het systeem te zijn en leidde tot een reeks structurele aanvallen.

De eerste grote doorbraak in de cryptanalyse werd in 1995 gerealiseerd door Gibson [34]. Hij toonde aan dat de structuur van de verstoringsmatrix X kon worden uitgebuit. Door de publieke matrix te manipuleren, kon hij het effect van de vervormingsmatrix elimineren en de publieke sleutel reduceren tot een stelsel van vergelijkingen dat in haalbare tijd kon worden opgelost. Als reactie hierop stelde Gabidulin verschillende varianten voor, waarbij extra vervormingsmatrices en meervoudige verstoringsmatrices werden toegevoegd om de onderliggende code beter te verbergen en Gibsons reducties onmogelijk te maken. Deze oplapmiddelen boden echter slechts schijnveiligheid.

In 2005 diende Overbeck [80] een fatale klap toe aan het GPT-systeem. Hij introduceerde een operator Λ_f die een matrix uitbreidt door er verticaal opeenvolgende Frobenius-machten aan toe te voegen. Overbeck buitte hiermee de fundamentele zwakte van Gabidulin-codes uit: wanneer Λ_f op de rijen van een Gabidulin code wordt toegepast, groeit de rang slechts lineair en uiterst traag,

omdat de Frobenius-machten de code op zichzelf afbeelden. Echter, wanneer Λ_f op willekeurige verstoringsmatrices wordt toegepast, groeit de rang exponentieel tot de volledige ruimte verzadigd is; zie ook Figuur 4.1. Hierdoor kon Overbeck de toegevoegde ruis wiskundig isoleren van het gestructureerde Gabidulin-gedeelte en in polynomiale tijd een equivalente geheime sleutel reconstrueren. Latere werken verfijnden deze resultaten en brachten de genadeklap toe aan de laatst mogelijke oplapmiddelen. De uiteindelijke conclusie is onverbiddelijk: klassieke Gabidulin-codes lekken inherent te veel informatie via hun algebraïsche structuur en zijn in de praktijk onveilig voor asymmetrische cryptografie.

Nieuwe MRD-codes via eindige meetkunde

Omdat de $\mathbb{F}_{q^m}$-lineariteit van Gabidulin-codes cryptografisch onveilig is gebleken, is de zoektocht naar nieuwe, minder gestructureerde ($\mathbb{F}_q$-lineaire) MRD-codes een van de belangrijkste open problemen in de code-gebaseerde cryptografie. In deze thesis tonen we aan hoe de eindige meetkunde hier een elegante oplossing voor biedt. Door de algebraïsche voorwaarden van rangmetrische codes te vertalen naar een meetkundig raamwerk, wordt de constructie van MRD-codes equivalent aan het vinden van *maximum h-scattered subspaces* en bijhorende *linear sets* in projectieve ruimten.

We beschrijven expliciet een constructiemethode uit de recente literatuur om rangmetrische codes te bouwen vanuit vectorruimten. Zij $V = V(r, q^m)$ een vectorruimte over $\mathbb{F}_{q^m}$, en $W = V(rm - k, q)$ een vectorruimte over het basisveld $\mathbb{F}_q$. Veronderstel dat $U = V(k, q)$ een $\mathbb{F}_q$-deelruimte is van V, en zij $G\colon V \to W$ een $\mathbb{F}_q$-lineaire afbeelding met kern $\ker(G) = U$. Voor elke $v \in V$ definiëren we de scalaire vermenigvuldiging in $\mathbb{F}_{q^m}$ als $\tau_v\colon \mathbb{F}_{q^m} \to V, \lambda \mapsto \lambda v$. Hiermee kunnen we een $\mathbb{F}_q$-lineaire rangmetrische code construeren als de verzameling van samenstellingen:

$$\mathcal{C}_{U,G} = \{G \circ \tau_v : v \in V\}.$$

De meetkundige eigenschappen van de deelruimte U bepalen rechtstreeks de parameters van de resulterende code $\mathcal{C}_{U,G}$. De cruciale invariant is hierbij de maximale dimensie van de doorsnede van U met één-dimensionale $\mathbb{F}_{q^m}$-deelruimten van V. Formeel definiëren we $\iota = \max\{\dim_{\mathbb{F}_q}(U \cap \langle v\rangle_{\mathbb{F}_{q^m}}) : v \in V \setminus \{0\}\}$. De volgende stelling stelt exact de meetkundige voorwaarden vast waaronder deze constructie een optimale MRD code oplevert.

Stelling (Theorem 3.4.3). *Zij $V = V(r, q^m)$ en $W = V(rm-k, q)$. Zij $U = V(k, q)$ een $\mathbb{F}_q$-deelruimte van V, $G\colon V \to W$ een $\mathbb{F}_q$-lineaire afbeelding met* $\ker(G) = U$*, en veronderstel dat $\iota < m$. Dan is de code $\mathcal{C}_{U,G}$ een $\mathbb{F}_q$-lineaire MRD code als en slechts als $\iota + 1$ een deler is van rm en de dimensie k voldoet aan*

$$k = \frac{\iota rm}{\iota + 1} \leq (r - 1)m.$$

Wanneer aan deze meetkundige grenzen is voldaan, bekomen we MRD-codes die, onder lichte voorwaarden, wiskundig bewezen niet-equivalent zijn aan eender welke variant van de Gabidulin-codes. Deze meetkundige bril verlegt de uitdaging van de cryptografie succesvol naar de eindige meetkunde: de zoektocht naar nieuwe veilige codes is nu herleid tot het construeren van *scattered subspaces*, al moeten er voor gebruik wel nog efficiënte decodeeralgoritmen voor dergelijke codes gevonden worden.

Deel II: Symmetrische cryptografie

Het tweede luik van deze masterproef bestudeert de veiligheidseigenschappen van kwadratische vectoriële functies $F\colon \mathbb{F}_p^n \to \mathbb{F}_p^m$ die in symmetrische blokcijfers dienst doen als substitutieboxen (S-boxen). De robuustheid van deze functies tegen cryptanalyse wordt gemeten aan de hand van het Walsh-spectrum, dat kwantificeert hoezeer componentfuncties $F_b = \langle b, F\rangle$ van F, met $b \in \mathbb{F}_p^m$, afwijken van zuiver affiene functies. Componenten met de kleinst mogelijke amplitude bieden optimale niet-lineariteit en worden gebogen (Eng.: *bent*) genoemd. Componenten die niet-gebogen zijn, vormen potentiële

cryptografische zwaktes. Het in kaart brengen van de verzameling van deze niet-gebogen componenten, genoteerd als N_F, is bijgevolg van fundamenteel belang.

In karakteristiek twee ($p = 2$) induceert de structuur van het Walsh-spectrum van kwadratische *almost perfect nonlinear* (APN) functies een vectorruimtepartitie (Eng.: *vector space partition*, VSP) op het domein. In recent werk toonden Bénéteau et al. [7] aan dat de verzameling N_F in karakteristiek twee een *blocking set* vormt in de projectieve ruimte $\mathrm{PG}(m-1, 2)$. In dit werk breiden we dit perspectief uit en karakteriseren we N_F als een projectief hyperoppervlak (Eng.: *projective hypersurface*). Zodoende lossen we een open probleem uit de literatuur op. We bewijzen expliciet dat de voorgestelde ondergrens voor de grootte van N_F in [7, Theorem 3.9] niet bereikt kan worden voor zogenaamde niet-MNBC (Eng.: *maximum number of bent components*) functies, waartoe APN-functies behoren in deze context.

Stelling (Theorem 5.4.10). *Zij n even met $n \leq 2m - 6$. Er bestaat* geen *niet-MNBC kwadratische functie $F \colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ met $|N_F| = 2^{m-n/2} + 2^{m-n/2-2} + 2^{m-n/2-3} - 1$.*

Vervolgens maken we gebruik van de gewichtsdistributie van Reed-Muller codes, verbonden via de veelterm die ons hyperoppervlak definieert, om een nieuwe en aantoonbaar scherpe ondergrens vast te leggen. Dit levert tevens een bovengrens op voor het aantal gebogen componenten van kwadratische APN-functies.

Stelling (Corollary 5.4.17). *Zij $n \geq 4$ even en $F \colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ een kwadratische APN-functie. Dan geldt*

$$|N_F| \geq 2^{n/2} + 2^{n/2-1} - 1.$$

Met andere woorden, het aantal gebogen componenten van F is kleiner dan of gelijk aan $2^n - 2^{n/2} + 2^{n/2-1}$.

Terwijl de theorie voor Booleaanse functies ($p = 2$) vrij uitgebreid is, blijven resultaten voor vectoriële functies over velden met oneven karakteristiek ($p > 2$) grotendeels uit. In deze thesis breiden we de theorie met succes uit naar velden van oneven karakteristiek via de klasse van *crooked* kwadratische functies. Hoewel het fundamentele concept van een vectorruimtepartitie naadloos behouden blijft, verliezen we de karakterisatie als *blocking set*. De onderliggende structuur van N_F vormt enkel nog een projectief hyperoppervlak, waarmee we ook een structurele beperking kunnen stellen op N_F.

Stelling (Theorem 5.4.18). *Zij $p > n \geq 2$ en zij $F \colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ een kwadratische functie. Beschouw een willekeurige rechte $\mathcal{L} \subseteq \mathrm{PG}(n-1, p)^*$. Zij $k = |\mathcal{L} \cap N_F|$ het aantal niet-gebogen componenten dat op deze rechte ligt. Dan moet k noodzakelijk voldoen aan één van de volgende voorwaarden:*

1. *$k \leq n$, of*
2. *$k = p + 1$ (de rechte is volledig bevat in het hyperoppervlak N_F).*

Dit tweede deel toont aan hoe het vertalen van algebraïsche en cryptografische eigenschappen naar de eindige meetkunde niet alleen hun analyse vereenvoudigt, maar ook sterkere voorwaarden stelt voor het ontwerpen van veilige symmetrische cijfers.

Besluit

Deze masterproef illustreert overtuigend dat eindige meetkunde ons kan helpen in cryptografisch onderzoek. In het domein van de asymmetrische cryptografie leidt de meetkundige vertaling tot een uitweg uit de kwetsbaarheden van de Gabidulin-codes, resulterend in methoden om nieuwe MRD-codes te construeren. In de symmetrische cryptografie leveren meetkundige objecten zoals *blocking sets* en projectieve hyperoppervlakken de geschikte structuur om theoretische grenzen aan te scherpen en het bestaan van bepaalde vectoriële functies uit te sluiten.

References

[1] Gianira N. Alfarano et al. 'Linear cutting blocking sets and minimal codes in the rank metric'. In: *Journal of Combinatorial Theory, Series A* 192 (2022), p. 105658.

[2] Simeon Ball. 'On sets of vectors of a finite vector space in which every subset of basis size is a basis'. In: *Journal of the European Mathematical Society* 14.3 (2012), pp. 733–748.

[3] Daniele Bartoli et al. 'Exceptional scattered sequences'. In: *Algebraic Combinatorics* 7.5 (2024), pp. 1405–1431.

[4] Daniele Bartoli et al. 'Maximum Scattered Linear Sets and Complete Caps in Galois Spaces'. In: *Combinatorica* 38.2 (2018), pp. 255–278.

[5] Hugo Beeloo-Sauerbier Couvée, Antonia Wachter-Zeh and Violetta Weger. 'Hybrid Subsupport Guessing: A New Hybrid Technique for the Rank Decoding Problem'. In: *International Conference on Post-Quantum Cryptography*. Springer. 2026, pp. 130–155.

[6] Thomas D. Bending and Dmitry Fon-Der-Flaass. 'Crooked Functions, Bent Functions, and Distance Regular Graphs'. In: *The Electronic Journal of Combinatorics* 5.1 (1998), R34.

[7] Sophie Hannah Bénéteau et al. 'On the Walsh spectra of quadratic APN functions'. In: *arXiv* (2025). arXiv: 2510.12008.

[8] Elwyn Berlekamp. 'Goppa codes'. In: *IEEE Transactions on Information Theory* 19.5 (1973), pp. 590–592.

[9] Elwyn Berlekamp, Robert McEliece and Henk Van Tilborg. 'On the inherent intractability of certain coding problems (corresp.)' In: *IEEE Transactions on Information Theory* 24.3 (2003), pp. 384–386.

[10] Daniel J. Bernstein. 'Grover vs. McEliece'. In: *International Workshop on Post-Quantum Cryptography*. Springer. 2010, pp. 73–80.

[11] Daniel J. Bernstein, Tanja Lange and Christiane Peters. 'Attacking and Defending the McEliece Cryptosystem'. In: *International Workshop on Post-Quantum Cryptography*. Springer. 2008, pp. 31–46.

[12] Claude Carlet. *Boolean Functions for Cryptography and Coding Theory*. Cambridge University Press, 2020. ISBN: 9781108606806.

[13] Claude Carlet, Pascale Charpin and Victor Zinoviev. 'Codes, Bent Functions and Permutations Suitable For DES-like Cryptosystems'. In: *Designs, Codes and Cryptography* 15.2 (1998), pp. 125–156.

[14] Arthur Cayley. 'Sur les déterminants gauches. (Suite du Mémoire T. XXXII. p. 119).' In: *Journal für die reine und angewandte Mathematik* 38 (1849), pp. 93–96.

[15] Pascale Charpin. 'The Crooked property'. In: *Finite Fields and Their Applications* 81 (2022), p. 102032.

[16] Alain Couvreur and Ilaria Zappatore. 'An Extension of Overbeck's Attack with an Application to Cryptanalysis of Twisted Gabidulin-Based Schemes'. In: *International Conference on Post-Quantum Cryptography*. Springer. 2023, pp. 3–37.

[17] Bence Csajbók and Alessandro Siciliano. 'Puncturing maximum rank distance codes'. In: *Journal of Algebraic Combinatorics* 49.4 (2019), pp. 507–534.

[18] Bence Csajbók et al. 'A new family of MRD-codes'. In: *Linear Algebra and its Applications* 548 (2018), pp. 203–220.

[19] Bence Csajbók et al. 'Generalising the Scattered Property of Subspaces'. In: *Combinatorica* 41.2 (2021), pp. 237–262.

[20] Bence Csajbók et al. 'Maximum scattered linear sets and MRD-codes'. In: *Journal of Algebraic Combinatorics* 46.3 (2017), pp. 517–531.

[21] Philippe Delsarte. 'Bilinear forms over a finite field, with applications to coding theory'. In: *Journal of Combinatorial Theory, Series A* 25.3 (1978), pp. 226–241.

[22] Peter Dembowski and Thedore G. Ostrom. 'Planes of order n with collineation groups of order n^2'. In: *Mathematische Zeitschrift* 103.3 (1968), pp. 239–258.

[23] Whitfield Diffie and Martin E. Hellman. 'New Directions in Cryptography'. In: *IEEE Transactions on Information Theory* 22.6 (1976), pp. 644–654.

[24] Lynn Engelberts, Simona Etinski and Johanna Loyer. 'Quantum sieving for code-based cryptanalysis and its limitations for ISD'. In: *Designs, Codes and Cryptography* (2025), pp. 1–34.

[25] Lance Fortnow. 'Fifty years of P vs. NP and the possibility of the impossible'. In: *Communications of the ACM* 65.1 (2021), pp. 76–85.

[26] Ernst M. Gabidulin. 'Attacks and counter-attacks on the GPT public key cryptosystem'. In: *Designs, Codes and Cryptography* 48.2 (2008), pp. 171–177.

[27] Ernst M. Gabidulin. *Public-key cryptosystems based on linear codes*. Delft University of Technology, Faculty of Technical Mathematics and Informatics, 1995.

[28] Ernst M. Gabidulin. 'Theory of Codes with Maximum Rank Distance'. In: *Problemy Peredachi Informatsii* 21.1 (1985), pp. 3–16.

[29] Ernst M. Gabidulin, A. V. Paramonov and O. V. Tretjakov. 'Ideals over a Non-Commutative Ring and their Application in Cryptology'. In: *Workshop on the Theory and Application of Cryptographic Techniques*. Springer. 1991, pp. 482–489.

[30] Ernst M. Gabidulin, Haitham Rashwan and Bahram Honary. 'On improving security of GPT cryptosystems'. In: *2009 IEEE International Symposium on Information Theory*. IEEE. 2009, pp. 1110–1114.

[31] Ernst M. Gabidulin et al. 'Reducible rank codes and their applications to cryptography'. In: *IEEE Transactions on Information Theory* 49.12 (2003), pp. 3289–3293.

[32] Philippe Gaborit and Gilles Zémor. 'On the hardness of the decoding and the minimum distance problems for rank codes'. In: *IEEE Transactions on Information Theory* 62.12 (2016), pp. 7245–7252.

[33] Philippe Gaborit et al. 'Low Rank Parity Check codes and their application to cryptography'. In: *Proceedings of the Workshop on Coding and Cryptography WCC*. 2013.

[34] Keith Gibson. 'Severely denting the Gabidulin version of the McEliece public key cryptosystem'. In: *Designs, Codes and Cryptography* 6.1 (1995), pp. 37–45.

[35] Keith Gibson. 'The Security of the Gabidulin Public Key Cryptosystem'. In: *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 1996, pp. 212–223.

[36] Robert Gold. 'Maximal recursive sequences with 3-valued recursive cross-correlation functions (Corresp.)' In: *IEEE transactions on Information Theory* 14.1 (1968), pp. 154–156.

[37] Patrick Govaerts and Leo Storme. 'The classification of the smallest nontrivial blocking sets in $PG(n, 2)$'. In: *Journal of Combinatorial Theory, Series A* 113.7 (2006), pp. 1543–1548.

[38] Rod Gow and Rachel Quinlan. 'Galois theory and linear algebra'. In: *Linear Algebra and its Applications* 430.7 (2009), pp. 1778–1789.

[39] Olof Heden. 'On the length of the tail of a vector space partition'. In: *Discrete Mathematics* 309.21 (2009), pp. 6169–6180.

[40] Olof Heden et al. 'Extremal sizes of subspace partitions'. In: *Designs, Codes and Cryptography* 64.3 (2012), pp. 265–274.

[41] James W. P. Hirschfeld. *Projective Geometries over Finite Fields*. Oxford University Press, 1998.

[42] Anna-Lena Horlemann-Trautmann, Kyle Marshall and Joachim Rosenthal. 'Extension of Overbeck's attack for Gabidulin-based cryptosystems'. In: *Designs, Codes and Cryptography* 86.2 (2018), pp. 319–340.

[43] William C. Huffman and Vera Pless. *Fundamentals of Error-Correcting Codes*. Cambridge University Press, 2010.

[44] Ghazal Kachigar and Jean-Pierre Tillich. 'Quantum Information Set Decoding Algorithms'. In: *International Workshop on Post-Quantum Cryptography*. Springer. 2017, pp. 69–89.

[45] Wrya K. Kadir and Chunlei Li. 'On decoding additive generalized twisted Gabidulin codes'. In: *Cryptography and Communications* 12.5 (2020), pp. 987–1009.

[46] Tadao Kasami and Nobuki Tokura. 'On the Weight Structure of Reed-Muller codes'. In: *IEEE Transactions on Information Theory* 16.6 (1970), pp. 752–759.

[47] Tadao Kasami, Nobuki Tokura and Saburo Azumi. 'On the Weight Enumeration of Weights Less than $2.5d$ of Reed—Muller Codes'. In: *Information and Control* 30.4 (1976), pp. 380–395.

[48] Neal Koblitz. 'Elliptic curve cryptosystems'. In: *Mathematics of computation* 48.177 (1987), pp. 203–209.

[49] Janne I. Kokkala, Denis S. Krotov and Patric R. J. Östergård. 'On the classification of MDS codes'. In: *IEEE Transactions on Information Theory* 61.12 (2015), pp. 6485–6492.

[50] Ralf Köotter and Frank R. Kschischang. 'Coding for Errors and Erasures in Random Network Coding'. In: *IEEE Transactions on Information Theory* 54.8 (2008), pp. 3579–3591.

[51] Alexander Kshevetskiy and Ernst Gabidulin. 'The new construction of rank codes'. In: *International Symposium on Information Theory*. IEEE. 2005, pp. 2105–2108.

[52] Serge Lang. *Algebra*. Graduate Texts in Mathematics. Springer New York, NY, 2002. ISBN: 978-0-387-95385-4.

[53] Juliane Lehmann and Olof Heden. 'Some necessary conditions for vector space partitions'. In: *Discrete Mathematics* 312.2 (2012), pp. 351–361.

[54] Yuan Xing Li, Robert H. Deng and Xin Mei Wang. 'On the equivalence of McEliece's and Niederreiter's public-key cryptosystems'. In: *IEEE Transactions on Information Theory* 40.1 (1994), pp. 271–273.

[55] Rudolf Lidl and Harald Niederreiter. *Finite Fields*. Vol. 20. Encyclopedia of Mathematics and its Applications. Reading, MA: Addison-Wesley, 1983.

[56] Dirk Liebhold and Gabriele Nebe. 'Automorphism groups of Gabidulin-like codes'. In: *Archiv der Mathematik* 107.4 (2016), pp. 355–366.

[57] Pierre Loidreau. 'A Welch–Berlekamp Like Algorithm for Decoding Gabidulin Codes'. In: *International Workshop on Coding and Cryptography*. Springer. 2005, pp. 36–45.

[58] Pierre Loidreau. 'Designing a Rank Metric Based McEliece Cryptosystem'. In: *International Workshop on Post-Quantum Cryptography*. Springer. 2010, pp. 142–152.

[59] Giovanni Longobardi. 'Scattered polynomials: an overview on their properties, connections and applications'. In: *arXiv* (2024). arXiv: 2411.11855.

[60] Giovanni Longobardi and Corrado Zanella. 'Linear sets and MRD-codes arising from a class of scattered linearized polynomials'. In: *Journal of Algebraic Combinatorics* 53.3 (2021), pp. 639–661.

[61] Guglielmo Lunardon. 'MRD-codes and linear sets'. In: *Journal of Combinatorial Theory, Series A* 149 (2017), pp. 1–20.

[62] Guglielmo Lunardon. 'Normal Spreads'. In: *Geometriae Dedicata* 75.3 (1999), pp. 245–261.

[63] Guglielmo Lunardon and Olga Polverino. 'Blocking Sets and Derivable Partial Spreads'. In: *Journal of Algebraic Combinatorics* 14.1 (2001), pp. 49–56.

[64] Guglielmo Lunardon and Olga Polverino. 'Translation ovoids of orthogonal polar spaces'. In: *Forum Mathematicum*. Vol. 16. 5. De Gruyter, 2004, pp. 663–669.

[65] Guglielmo Lunardon, Rocco Trombetti and Yue Zhou. 'Generalized twisted Gabidulin codes'. In: *Journal of Combinatorial Theory, Series A* 159 (2018), pp. 79–106.

[66] Guglielmo Lunardon, Rocco Trombetti and Yue Zhou. 'On kernels and nuclei of rank metric codes'. In: *Journal of Algebraic Combinatorics* 46.2 (2017), pp. 313–340.

[67] Giuseppe Marino, Alessandro Neri and Rocco Trombetti. 'Evasive subspaces, generalized rank weights and near MRD codes'. In: *Discrete Mathematics* 346.12 (2023), p. 113605.

[68] Robert J. McEliece. 'A public-key cryptosystem based on algebraic coding theory'. In: *DSN Progress Report* 42-44 (1978), pp. 114–116.

[69] Sihem Mesnager et al. 'Further study on the maximum number of bent components of vectorial functions'. In: *Designs, Codes and Cryptography* 87.11 (2019), pp. 2597–2610.

[70] Victor S. Miller. 'Use of elliptic curves in cryptography'. In: *Conference on the theory and application of cryptographic techniques*. Springer. 1985, pp. 417–426.

[71] Esmeralda Năstase and Papa Sissokho. 'The second minimum size of a finite subspace partition'. In: *Designs, Codes and Cryptography* 94.3 (2026), p. 53.

[72] Harald Niederreiter. 'Knapsack-type cryptosystems and algebraic coding theory'. In: *Problems of Control and Information Theory* 15.2 (1986), pp. 157–166.

[73] Kaisa Nyberg. 'Perfect nonlinear S-boxes'. In: *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer. 1991, pp. 378–386.

[74] Open Quantum Safe. *Classic McEliece*. URL: https://openquantumsafe.org/liboqs/algorithms/kem/classic_mceliece.html (visited on 20/01/2026).

[75] Kamil Otal and Ferruh Özbudak. 'Additive rank metric codes'. In: *IEEE Transactions on Information Theory* 63.1 (2016), pp. 164–168.

[76] Kamil Otal and Ferruh Özbudak. 'Explicit constructions of some non-Gabidulin linear maximum rank distance codes'. In: *Advances in Mathematics of Communications* 10.3 (2016).

[77] Ayoub Otmani, Hervé Talé Kalachi and Sélestin Ndjeya. 'Improved cryptanalysis of rank metric schemes based on Gabidulin codes'. In: *Designs, Codes and Cryptography* 86.9 (2018), pp. 1983–1996.

[78] Alexei V. Ourivski and Ernst M. Gabidulin. 'Column scrambler for the GPT cryptosystem'. In: *Discrete Applied Mathematics* 128.1 (2003), pp. 207–221.

[79] Raphael Overbeck. 'A New Structural Attack for GPT and Variants'. In: *International Conference on Cryptology in Malaysia*. Springer. 2005, pp. 50–63.

[80] Raphael Overbeck. 'Extending Gibson's Attacks on the GPT Cryptosystem'. In: *International Workshop on Coding and Cryptography*. Springer. 2005, pp. 178–188.

[81] Raphael Overbeck. 'Structural Attacks for Public Key Cryptosystems based on Gabidulin Codes'. In: *Journal of cryptology* 21.2 (2008), pp. 280–301.

[82] Olga Polverino. 'Linear sets in finite projective spaces'. In: *Discrete mathematics* 310.22 (2010), pp. 3096–3107.

[83] Olga Polverino and Ferdinando Zullo. 'Connections between scattered linear sets and MRD-codes'. In: *Bull. Inst. Combin. Appl* 89 (2020), pp. 46–74.

[84] Alexander Pott et al. 'On the Maximum Number of Bent Components of Vectorial Functions'. In: *IEEE Transactions on Information Theory* 64.1 (2017), pp. 403–411.

[85] Eugene Prange. 'The use of information sets in decoding cyclic codes'. In: *IRE Transactions on Information Theory* 8.5 (1962), pp. 5–9.

[86] Tovohery Hajatiana Randrianarisoa. 'A decoding algorithm for rank metric codes'. In: *arXiv* (2017). arXiv: 1712.07060.

[87] Tovohery Hajatiana Randrianarisoa. 'A geometric approach to rank metric codes and a classification of constant weight codes'. In: *Designs, Codes and Cryptography* 88.7 (2020), pp. 1331–1348.

[88] Haitham Rashwan, Ernst M. Gabidulin and Bahram Honary. 'A Smart approach for GPT cryptosystem based on rank codes'. In: *2010 IEEE International Symposium on Information Theory*. IEEE. 2010, pp. 2463–2467.

[89] Alberto Ravagnani. 'Rank-metric codes and their duality theory'. In: *Designs, Codes and Cryptography* 80.1 (2016), pp. 197–216.

[90] Ronald L. Rivest, Adi Shamir and Leonard Adleman. 'A method for obtaining digital signatures and public-key cryptosystems'. In: *Communications of the ACM* 21.2 (1978), pp. 120–126.

[91] Paolo Santonastaso. 'Algebraic and Geometric Methods in Rank-metric Codes'. PhD Thesis. Università degli Studi della Campania "Luigi Vanvitelli", Dec. 2023.

[92] Beniamino Segre. 'Teoria di Galois, fibrazioni proiettive e geometrie non desarguesiane'. In: *Annali di Matematica Pura ed Applicata* 64.1 (1964), pp. 1–76.

[93] Claude E. Shannon. 'Communication theory of secrecy systems'. In: *The Bell System Technical Journal* 28.4 (1949), pp. 656–715.

[94] John Sheekey. 'A new family of linear maximum rank distance codes'. In: *Advances in Mathematics of Communications* 10.3 (2016), pp. 475–488.

[95] John Sheekey. 'MRD codes: constructions and connections'. In: *Combinatorics and Finite Fields: Difference Sets, Polynomials, Pseudorandomness and Applications*. Ed. by Kai-Uwe Schmidt and Arne Winterhof. Berlin, Boston: De Gruyter, 2019, pp. 255–286. ISBN: 9783110642094.

[96] John Sheekey and Geertrui Van de Voorde. 'Rank-metric codes, linear sets, and their duality'. In: *Designs, Codes and Cryptography* 88.4 (2020), pp. 655–675.

[97] Peter W. Shor. 'Algorithms for quantum computation: discrete logarithms and factoring'. In: *Proceedings 35th Annual Symposium on Foundations of Computer Science*. IEEE. Santa Fe, NM, USA, 1994, pp. 124–134.

[98] Danilo Silva, Frank R. Kschischang and Ralf Kötter. 'A Rank-Metric Approach to Error Control in Random Network Coding'. In: *IEEE Transactions on Information Theory* 54.9 (2008), pp. 3951–3967.

[99] Richard C. Singleton. 'Maximum distance q-nary codes'. In: *IEEE Transactions on Information Theory* 10.2 (1964), pp. 116–118.

[100] Rocco Trombetti and Yue Zhou. 'A New Family of MRD Codes in $\mathbb{F}_q^{2n\times 2n}$ With Right and Middle Nuclei $\mathbb{F}_{q^n}$'. In: *IEEE Transactions on Information Theory* 65.2 (2018), pp. 1054–1062.

[101] Rocco Trombetti and Yue Zhou. 'Nuclei and automorphism groups of generalized twisted Gabidulin codes'. In: *Linear Algebra and its Applications* 575 (2019), pp. 1–26.

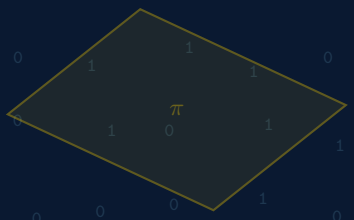
[102] Geertrui Van de Voorde. *Linear sets*. Lecture notes on Finite Geometry and Friends. 2019.

[103] Violetta Weger, Niklas Gassner and Joachim Rosenthal. 'A survey on code-based cryptography'. In: *arXiv* (2022). arXiv: 2201.07119.

[104] Giovanni Zini and Ferdinando Zullo. 'Scattered subspaces and related codes'. In: *Designs, Codes and Cryptography* 89.8 (2021), pp. 1853–1873.

projective hypersurface

$\Lambda_f(\mathcal{G}_{k,\sigma})$

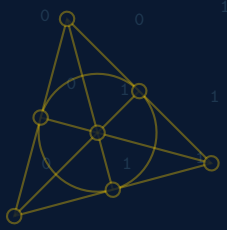


$$W_F(b,a) = \sum_{x \in \mathbb{F}_p^n} \xi_p^{F_b(x) - \langle a,x \rangle}$$

$C_{U,G}$

$F : \mathbb{F}_p^n \to \mathbb{F}_p^m$

$B \cap L \neq \emptyset$



$\mathbb{F}_{q^m}$

h-scattered subspace

GHENT UNIVERSITY

FACULTY OF SCIENCES