A study of cryptographic systems based on Rank metric codes

Marshall, Kyle Daniel

Abstract: The ubiquity, dependability, and extensiveness of internet access has seen a migration of local services to cloud services where the advantages of scalability can be efficiently exploited. In doing so, the exposure of sensitive data to eavesdropping is a principal concern. Asymmetric cryptosystems attempt to solve this problem by basing access on the knowledge of a solution to mathematically difficult problems. Shor demonstrated that on a quantum computer, cryptosystems based on the difficulty of factoring integers or solving discrete logarithms were efficiently solvable. As the most ubiquitous asymmetric cryptosystems in modern use are based on these problems, new cryptosystems had to be considered for post-quantum cryptography. In 1978, McEliece proposed a cryptosystem based on the difficulty of decoding random linear codes but the key sizes were too large for practical consideration. These systems, though, do appear to resist Shor's algorithm and other quantum attacks. More recently, Gabidulin proposed using codes in the rank metric to design secure cryptosystems because they could be designed with smaller parameters. In this direction, many proposals for cryptosystems based on rank metric codes were designed. Overbeck managed to cryptanalyze many of these systems, but there remain several which resist all known structural attacks. In this work, we investigate the use of rank metric codes for cryptographic purposes. Firstly, we investigate the construction of MRD codes and propose some new constructions based on combinatorial methods. We then generalize Overbeck's attack and show how our generalized attack can be used to cryptanalyze some of the cryptosystems which were designed to resist the attack of Overbeck. Our attack is based on a new approach of exploiting the structure of low weight elements in the code. Our approach also allows us to extend a result of Gaborit to obtain a polynomial time decoding algorithm for codes with certain parameters. Lastly, we consider the use of codes in the subspace metric– which are based on rank metric codes–in order to create an alternative instance of Juels' and Sudan's fuzzy vault primitive.

A Study of Cryptographic Systems based on Rank Metric Codes

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich
von

Kyle Daniel Marshall

aus

den USA

Promotionskomitee
Prof. Dr. Joachim Rosenthal (Vorsitz)
Prof. Dr. Andrew Kresch (Universität Zürich)
Dr. Anna-Lena Horlemann-Trautmann (EPFL)

Externe Gutachter
Prof. Dr. Nigel Boston (University of Wisconsin - Madison)
Prof. Dr. Elisa Gorla (Université de Neuchâtel)

Zürich, 2016

---

to my family and friends

Acknowledgements

Foremost, I would like to thank my friends and family for their constant support. For those who traveled such a long way to bring a piece of home to me when I could not return, I am especially grateful. I consider myself incredibly fortunate to have so many beautiful relationships.

I would also like to thank my advisor, Prof. Joachim Rosenthal, for supporting my research interests and allowing me the freedom to explore my own interests and ideas. His direction and advice have been invaluable for me during my studies. I would also like to thank all my coauthors and collaborators–Dr. Michael O'Sullivan, Dr. Fernando Hernando, Dr. Felix Fontein, and Dr. Davide Schipani–with special recognition of Dr. Anna-Lena Trautmann-Horlemann. Her expertise, guidance, and enthusiasm have been crucial to the development of the ideas in this thesis and have helped me to become a more effective researcher.

Lastly, I would like to thank the University of Zürich and especially the Institute für Mathematik for the opportunities provided during my doctoral studies as well as the Swiss National Science Foundation for their financial support during my studies.

Abstract

The ubiquity, dependability, and extensiveness of internet access has seen a migration of local services to cloud services where the advantages of scalability can be efficiently exploited. In doing so, the exposure of sensitive data to eavesdropping is a principal concern. Asymmetric cryptosystems attempt to solve this problem by basing access on the knowledge of a solution to mathematically difficult problems. Shor demonstrated that on a quantum computer, cryptosystems based on the difficulty of factoring integers or solving discrete logarithms were efficiently solvable. As the most ubiquitous asymmetric cryptosystems in modern use are based on these problems, new cryptosystems had to be considered for post-quantum cryptography. In 1978, McEliece proposed a cryptosystem based on the difficulty of decoding random linear codes but the key sizes were too large for practical consideration. These systems, though, do appear to resist Shor's algorithm and other quantum attacks. More recently, Gabidulin proposed using codes in the rank metric to design secure cryptosystems because they could be designed with smaller parameters. In this direction, many proposals for cryptosystems based on rank metric codes were designed. Overbeck managed to cryptanalyze many of these systems, but there remain several which resist all known structural attacks.

In this work, we investigate the use of rank metric codes for cryptographic purposes. Firstly, we investigate the construction of MRD codes and propose some new constructions based on combinatorial methods. We then generalize Overbeck's attack and show how our generalized attack can be used to cryptanalyze some of the cryptosystems which were designed to resist the attack of Overbeck. Our attack is based on a new approach of exploiting the structure of low weight elements in the code. Our approach also allows us to extend a result of Gaborit to obtain a polynomial time decoding algorithm for codes with certain parameters. Lastly, we consider the use of codes in the subspace metric– which are based on rank metric codes–in order to create an alternative instance of Juels' and Sudan's fuzzy vault primitive.

Zusammenfassung

Die Allgegenwart, Zuverlässigkeit und Weitläufigkeit des Internet-Zuganges erlebt eine Migration von lokalen Diensten zu Dienste auf der Cloud, wo die Vorteile der Skalierbarkeit effizient genutzt werden können. Dabei ist die Gefährdung von sensiblen Daten durch Lauschangriffen ein Hauptanliegen. Asymmetrische Kryptosysteme versuchen, dieses Problem zu lösen, indem Zugriff auf Daten, basierend auf der Lösung eines mathematisch schwierigen Problem, vergeben wird. Shor zeigte, dass auf einem Quantencomputer, auf der Schwierigkeit natürliche Zahlen zu faktorisieren oder Lösungen diskreter Logarithmen basierende Kryptosysteme, effizient lösbar sind. Da die am weitesten verbreiteten asymmetrischen Kryptosysteme heutzutage auf diesen Problemen beruhen, mussten neue Kryptosysteme für die Post-Quantenkryptographie berücksichtigt werden. McEliece hat ein Verschlüsselungssystem basierend auf der Schwierigkeit "random linear codes" zu entschlüsseln, aber die Schlüsselgrößen waren zu groß für praktische Anwendungen. Dieses System jedoch widersteht Shor's Algorithmus und andere Quanten Angriffen. Neuerdings hat Gabidulin Codes in der Rangmetrik vorgeschlagen um sichere Kryptosysteme zu entwerfen, weil sie mit kleineren Parametern entwickelt werden könnten. Danach wurden viele Vorschläge für Kryptosysteme auf Basis von Rangmetrik-Codes entworfen. Overbeck konnte viele dieser Systeme Kryptanalysieren, aber es bleiben einige, die allen bekannten strukturellen Angriffen widerstehen.

In dieser Arbeit untersuchen wir die Verwendung von Rangmetrik-codes für kryptographische Zwecke. Zunächst untersuchen wir den Bau von MRD-codes und präsentieren einige neue Konstruktionen basierend auf kombinatorische Verfahren. Wir verallgemeinern dann Overbecks Angriff und zeigen, wie unser allgemeiner Angriff verwendet werden kann um einige der Kryptosysteme zu Kryptanalysieren, die so entworfen wurden, dass sie den Angriff von Overbeck widerstehen. Unser Angriff basiert auf die Nutzung der Struktur von Elementen im Code mit geringem Gewicht. Unser Ansatz erlaubt es uns auch ein Polynomialer Zeit -Decodieralgorithmus für Codes mit bestimmten Parametern zu erhalten, was die Erweiterung eines Ergebnis von Gaborit ist. Schließlich betrachten wir die Verwendung von Codes in der Unterraummetrik -, die auf Rangmetrik-Codes basieren - um eine alternative Instanz von "Juels' and Sudan's fuzzy vault primitive" zu erstellen.

Contents

Chapter 1

Foreword

Reliable and secure two-party communication is fundamental to the information age. The prevalence of distributed information systems–for instance cloud storage, e-banking, and social media–is in large part due to confidence in the ability to exchange information in an effective way. Two principal goals in this direction are reliability and security of the communication system. Reliability involves the ability of a receiving party to correctly observe an intended message whereas security refers to the ability for parties to communicate in secret. The former of these concerns was pioneered in the seminal work of Shannon in his 1948 paper entitled, *The Mathematical Theory of Communication* [69]. Considered one of the most influential works of the 20th century, the increase in computational power of modern computers has been followed by an increased interest in the applications of his work. His ideas can now ubiquitously be found in all communication systems, but the development of the field of information theory in the pre-computational era is a testament to his vision.

As an example of unreliable communication, consider a sender who wishes to transmit either a 1 or 0 to the receiver. However, in the process of communication, the message is corrupted in such a way that a uniformly random bit is received, regardless of the bit that was sent. That is,

$$P(x \text{ sent} \mid y \text{ received}) = .5,$$

for any $x, y \in \{0, 1\}$. The receiver in this case has no ability to discern if the observation of a 1 corresponded to a transmitted 0, or vice versa. In this case, there is no communication scheme which can allow the receiver to recover information about the sent message. If, on the other hand, we take $p < 1/2$ and we have

$$P(x \text{ sent} \mid y \text{ received}) = \begin{cases} p & x \neq y \\ 1-p & x = y, \end{cases}$$

then we have a better chance of assuming the received bit is correct than incorrect. We can use this information to our advantage; *encode* a bit x into a sequence of 3 bits, by repeating x three times. For instance, instead of just sending a 1, send 111. Then, if $y_1y_2y_3$ is received, at least two of y_1, y_2, or y_3 are equal to, $b \in \{0, 1\}$. Then, a receiver observing $y_1y_2y_3$ and guessing that b was transmitted has a good chance of being correct.

Specifically, the probability of incorrectly recovering the transmitted bit by this majority decoding scheme is

$$\sum_{i=2}^{3} \binom{3}{i} p^i (1-p)^{3-i}.$$

This value decreases to 0 as $p \to 0$. We can also, instead of repeating a bit 3 times, repeat it $2n+1$ times and then the probability of failure to recover b will be

$$\sum_{i=n+1}^{2n+1} \binom{2n+1}{i} p^i (1-p)^{2n+1-i}.$$

For any $p < 1/2$, one can choose n large enough that this value can be made arbitrarily small. Therefore, reliable communication can be achieved. The cost of being able to recover the message correctly is that the rate of information transmission must decrease. Out of every $2n+1$ symbols sent in the example above, only one message symbol can be transmitted. In general, this is not the most efficient encoding scheme, and the limits of such schemes are one of the fundamental results of Shannon [69].

The basic point-to-point communication model involves only one sender and one receiver, summarized in Figure 2.1. In modern communication systems, information is often distributed between different sources and requested by multiple receivers. The channel then consists of a network of nodes which transmit the information between each other before finally arriving at the intended receivers. The point-to-point model under these circumstances is no longer the most efficient means of communication. The study of networks grew out of a motivation to understand the limits of communication through networks, in which the sender(s) and receiver(s) are connected by nodes. While classical coding theory focuses on designing codes in the Hamming metric, the rank metric and subspace metric were found to be more amenable to the types of errors which occur in this network context. One of the fundamental results in this field came from Ahlswede et al. in which it was shown that by allowing the nodes to use algebraic operations rather than simple routing, the capacity of a random network could be attained [2]. Sophisticated techniques and constructions have since been developed, leading to a rich understanding of the subject and a wide variety of applications. Peer-to-peer systems, online gaming, distributed storage, and file-streaming services are examples in which communication can be improved by using network ideas [19, 38].

The latter of our considerations–security–has found significant traction as the storage of valuable data is increasing being handled remotely, where the need for authorized access is paramount. This has motivated a deeper understanding of complexity–the ability to describe how difficult certain problems are. While much remains to be discovered in this area and important questions remain, the understanding of complexity has provided a basis for which all secure communication is based. In order to safely use remote services, there must be a degree of confidence that the information being stored is secure from possibly nefarious parties. Obvious examples include remote access to online banking, social media accounts, or cloud storage services just to name a few. In order to communicate securely, the sender and receiver should communicate via a private key

which only they know, and which allows them to encrypt and decrypt information. To an observer, the encrypted messages should leak as little information as possible about the message itself and moreover the algorithm for encryption/decryption should be efficient. For remote parties without physical access or a secure channel, some method must be devised in order to share the private key. Obviously, the private key cannot be directly communicated, else an adversary could intercept it and also be able to decrypt. The study of public-key cryptosystems grew out of this problem and is a widely heavily studied topic due to its importance and applicability.

The most ubiqitous public-key cryptosystems are the RSA and ECC cryptosystems. Both are based on number theoretic problems which are considered difficult to solve in the absence of side information. The problem with both systems, however, is that they are very efficiently broken by a large enough quantum computer [71]. It is believed by some that the momentum of progress in quantum computing will inevitably result in a practical quantum computer, at which time cryptosystems systems based on factoring or the discrete logarithm will become vulnerable to attack. In the event of such a scenario, it is imperative that new types of public-key cryptosystems are properly analyzed to ensure they remain safe in the quantum computing age. In this direction, there are several candidates. One of the first proposed is the McEliece cryptosystem, based on the difficulty of decoding a random code. Lattice-based cryptosystems share a similar fundamental problem and, recently, rank-based cryptosystems have received some special attention. Considered very promising at first, many proposals were made and studied. In [34] Gibson gave the first reasonable attack on proposed rank-based cryptosystems and in [62] Overbeck extended the attack to break most of the existing proposals in the literature. The weakness of these proposals was their reliance on Gabidulin codes–highly structured codes in the rank metric. Since then, variants have been designed specifically to resist Overbeck's attack. Some are still based on Gabidulin codes, and some use entirely new families of codes [29, 49, 64, 31].

One of the impediments for using rank-based codes is the scarcity of known interesting families of decodable codes; it is why most early proposals used Gabidulin codes and, consequently, were efficiently attacked by Overbeck. Of particular interest amongst rank metric codes are the codes which meet the Singleton bound, called maximum rank distance (MRD) codes. These codes have optimal trade-off between the rate of communication and the proportion of correctable errors. The study of MRD codes has led to interesting constructions (for instance, see [70, 50, 16, 17]), although most are linear over a subfield of the alphabet. Codes which are linear over an extension field are of particular interest because they have a more efficient matrix representation, thus allowing for more efficient storage and consequently a reduction in the public key size. However, aside from a special case of generalized twisted Gabidulin codes, there have been no formal constructions for MRD codes which are linear over the extension field.

In Chapter 2 we present the necessary background for motivating the results of this thesis. In Chapter 3, we give some important constructions of codes in the rank metric, as well as properties of MRD codes. Using these properties, we can construct small examples of MRD codes which are linear over the extension field and are different from

generalized Gabidulin codes. These are new MRD codes which are linear as codes over the extension field and which are not equivalent to a generalized Gabidulin code. In order to show this, we also give an algebraic criterion for determining if a code is a generalized Gabidulin code. Chapter 4 contains a brief introduction to coding-based cryptography and its extension to the rank metric. Several variants of cryptosystems are presented, one of which has already been thoroughly cryptanalyzed by Overbeck–the GPT cryptosystem–as well as several variants that have up to now not been broken. In Chapter 5, we give a new attack that allows us to extend Overbeck's attack to break all the variants designed to resist it. This new attack is based on the consideration of a different notion of support than is taken in other approaches. We also apply this attack to a variant of the GPT designed by scrambling the private generator matrix by a non-isometry. In Chapter 6, we first present a cryptographic primitive based on imperfect key submission, first proposed by Juels and Sudan in [42]. We extend their idea to use codes in the subspace metric, rather than the Hamming metric and examine some security and design concerns.

Chapter 2

Background

One of the many insights of Shannon was to dissect the process of sending and receiving a message into modular processes in such a way that each of these modular processes can be examined independently. In order to understand these processes, he established much of the language of what would eventually become information theory. While any thorough treatment of the subject should include such language, we will not require such generality. Rather, in this thesis we consider only algebraic and combinatorial aspects in a restricted framework.

The point-to-point communication model we consider is presented in Figure 2.1. Our model differs slightly from the one presented in [54]. This communication structure is the impetus for the study of forward error correction; techniques that allow the receiver to recover the message in the presence of channel errors, without the need to communicate with the source. While the concept may appear simple, the techniques used can be quite sophisticated. Depending on the channel and other considerations regarding efficiency and reliability, codes can be quite exotic objects. All proofs and more details can be found in [54, 9]

Let $\mathbb{K}$ be a finite field, $N, n \in \mathbb{N}$. We mean by an *encoding* an injective map $\iota\colon \{0, \ldots, N-1\} \to \mathbb{K}^n$. The image of ι is called the *code*. The elements of a code are called *codewords*. Suppose that $\mathbb{K} = \mathbb{F}_q$ is the finite field of q elements and $N = q^k$ for $k < n$. Then, we can associate elements from $\{0, \ldots, q^k - 1\}$ with the elements of $\mathbb{K}^k$ and view ι as a map from $\mathbb{K}^k$ to $\mathbb{K}^n$. The image of an encoding map $\iota\colon \mathbb{K}^k \to \mathbb{K}^n$ which is linear will be called a *linear code*. We are primarily concerned with these codes for practical purposes. If the encoding map $\iota\colon \mathbb{K}^k \to \mathbb{K}^n$ is linear, and $\mathcal{C} = \iota(\mathbb{K}^k)$, then $\mathcal{C}$ can always be given as the image of a matrix $G \in \mathbb{K}^{k\times n}$. Any such matrix representing ι will be called a *generator matrix* for $\mathcal{C}$. Moreover, all generator matrices are equivalent up to multiplication on the left by an element of $\mathrm{GL}_k(\mathbb{K})$.

The channel is the medium through which a codeword must travel before it is observed by the receiver. During transmission, errors can occur, which will be modeled by a stochastic function, ρ, on $\mathbb{K}^n$. For a codeword $\boldsymbol{x} \in \mathbb{K}^n$, The decoder receives the *senseword*, $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}$, where $\boldsymbol{e}$ is a random element of $\mathbb{K}^n$ according to the distribution of ρ. $\boldsymbol{e}$ will be called the *error vector* associated to the senseword $\boldsymbol{y}$. The objective of forward

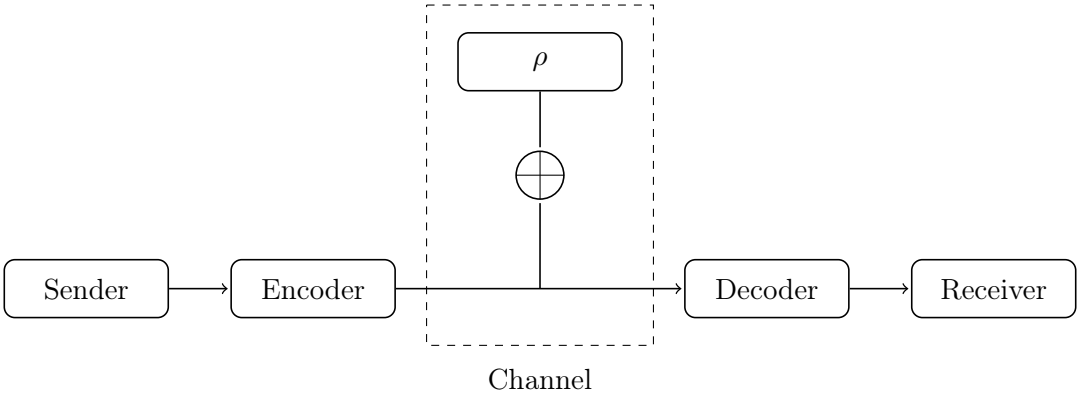

Figure 2.1: Point-to-Point Communication Model

error correction is to design the map ι in a clever way so that even in the presence of an error, the receiver can determine with high probability which codeword was sent. For linear codes, knowledge of the codeword allows the receiver to easily recover the message.

For a vector $\boldsymbol{x} \in \mathbb{K}^n$, denote by x_i the ith coordinate of $\boldsymbol{x}$. We can consider $\mathbb{K}^n$ as a metric space in the following way.

Definition 2.0.1. The Hamming metric, $d_{\mathrm{H}} \colon \mathbb{K}^n \times \mathbb{K}^n \to \{0, ..., n\}$ is given by

$$d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y}) = |\{i \mid x_i \neq y_i\}|.$$

i.e., the number of coordinates in which $\boldsymbol{x}$ and $\boldsymbol{y}$ differ.

Definition 2.0.2. The *Hamming weight* of a vector, $\boldsymbol{x}$ will be given by

$$\mathrm{wt}_{\mathrm{H}}(\boldsymbol{x}) = d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{0}),$$

i.e., the number of non-zero coordinates of $\boldsymbol{x}$.

Definition 2.0.3. Let $\mathcal{C} \subset \mathbb{K}^n$ be a code. The *minimum distance* of $\mathcal{C}$ is given by

$$d_{\mathrm{H}}^{\min}(\mathcal{C}) = \min\{d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}, \boldsymbol{x} \neq \boldsymbol{y}\}.$$

Definition 2.0.4. Let $\mathcal{C}$ be a k-dimensional subspace of $\mathbb{K}^n$ of minimum distance d. Then, $\mathcal{C}$ will be called an $[n, k, d]_{\mathbb{K}}$ linear code. The value n will be called the *length* of $\mathcal{C}$.

Definition 2.0.5. Let $\mathcal{C}$ be a linear code. The dual of $\mathcal{C}$, denoted $\mathcal{C}^{\perp}$, is given by

$$\mathcal{C}^{\perp} = \{\boldsymbol{y} \mid \boldsymbol{x}\boldsymbol{y}^T = 0, \text{ for all } \boldsymbol{x} \in \mathcal{C}\}.$$

The following theorem can be found on page 25 of [9].

Theorem 2.0.6. *Let $\mathcal{C}$ be a linear code of length n and dimension k. Then, $\mathcal{C}^{\perp}$ is a code of length n and dimension $n - k$.*

We can define the Hamming ball around $\boldsymbol{x} \in \mathbb{K}^n$ of radius r by

$$B^{\mathrm{H}}_{r,n}(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{K}^n \mid d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y}) \leq r\}.$$

Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ be a codeword of $\mathcal{C}$. Suppose now that $\mathbb{K} = \mathbb{F}_q$ is the finite field with cardinality q. Consider the channel which, for each coordinate x_i of $\boldsymbol{x}$, changes the value of x_i with probability p to a different element of $\mathbb{F}_q$ in a uniform way. Specifically, define ρ by

$$\rho(x_i) = x_i + \varepsilon,$$

where

$$\Pr(\varepsilon = a) = \begin{cases} 1-p & a = 0 \\ \dfrac{1}{q-1}p & a \in \mathbb{F}_q^*. \end{cases}$$

If we extend this to each coordinate of the vector, $\boldsymbol{x}$, then a receiver will obtain a vector of the form,

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e},$$

where the coordinates of $\boldsymbol{e}$ are determined by ρ. This channel is called the q-ary symmetric channel. It is natural that the Hamming distance allows us to correct errors in this metric. In particular, we have the following:

Lemma 2.0.7. *Let $\mathcal{C}$ be an $[n,k,d]_{\mathbb{F}_q}$ code in the Hamming metric. Let $\boldsymbol{x} \in \mathbb{F}_q^n$ be sent through the q-ary symmetric channel with transition probability $p < 1/2$ and $\boldsymbol{y}$ be received. Then, there is a unique $\boldsymbol{x}$ such that*

$$d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y}) \leq \left\lfloor \frac{d-1}{2} \right\rfloor$$

with probability

$$\sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{i} (1-p)^{n-i} p^i.$$

We note that if we fix n, k, and d, this value approaches 1 as $p \to 0$. On the other hand, if p is fixed, then there exists a code such that the probability of error in decoding is arbitrarily small. In general, the ability to create a code for a given channel is a function of the *capacity* of the channel. A more precise statement can be found in Chapter 2.2 of [54]. From the notion of the channel to that of a metric is a way of translating the problem into one of geometry. One can consider the problem as one of sphere packing, not necessarily arising from a channel. Different geometric problems are more amenable to different channels, but generally, for any metric d^*, we have,

Lemma 2.0.8. *Let $\mathcal{C}$ be an $[n,k,d]_{\mathbb{F}_q}$ code with respect to the metric d^*. Then, $\mathcal{C}$ can be used to correct any error vector, $\boldsymbol{e}$ satisfying $d^*(\boldsymbol{x}, \boldsymbol{x}+\boldsymbol{e}) \leq \lfloor \frac{d-1}{2} \rfloor$.*

There are different notions of correctability. What we mean here is that for any $\boldsymbol{y} \in \mathbb{F}_q^n$, there is *at most* one codeword, $\boldsymbol{x} \in \mathcal{C}$ such that $d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y}) \leq \left\lfloor \frac{d-1}{2} \right\rfloor$. Henceforth, when discussing error-correcting codes we will often do so without mention of a channel. One can assume, for instance, that there is an underlying channel producing errors which are compatible with the respective metric, or simply ignore the channel in the communication system altogether and consider the geometric packing problem as given in Lemma 2.0.8. In terms of packing, there is a clear trade-off between the radius of the spheres and the amount of spheres that can be packed into the space. An elementary estimate in the case of the Hamming metric is the well-known Singleton bound, a proof of which can be found on page 82 of [9].

Theorem 2.0.9 (Singleton)**.** *Let $\mathcal{C} \subset \mathbb{F}_q^n$. Then,*

$$d_{\mathrm{H}}^{\min}(\mathcal{C}) \leq n - \log_q |\mathcal{C}| + 1.$$

A code meeting the Singleton bound in the Hamming metric is called a maximum distance separable (MDS) code. While the Singleton bound is not achievable for all combinations of n and q, for $n < q$, there are examples of constructions for codes which are MDS. Perhaps the most ubiquitous is the following.

Definition 2.0.10. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ be a vector of distinct elements of $\mathbb{F}_q^*$. The Reed-Solomon code of length n and dimension k with evaluation vector $\boldsymbol{\alpha}$ is given by,

$$\mathrm{RS}_{n,k}(\boldsymbol{\alpha}) = \{(f(\alpha_1), \ldots, f(\alpha_n)) \mid \deg(f) < k\}.$$

Proposition 2.0.11. *Reed-Solomon codes are always MDS codes, that is,*

$$d_{\mathrm{H}}^{\min}(\mathrm{RS}_{n,k}(\boldsymbol{\alpha})) = n - k + 1.$$

The last step in Figure 2.1 is the decoding step. In Lemma 2.0.8, the decoding problem we consider is one of minimum distance decoding. In general, decoding is extremely difficult. In Chapter 4 the difficulty of this problem is discussed. However, for codes which are designed with some structure, efficient decoding algorithms can be given. One of the breakthrough results in coding theory in recent years was the efficient *list-decoding* of Reed-Solomon codes–the ability to compute all codewords within a ball of radius larger than half the minimum distance [37]. Other codes which have efficient decoding algorithms are BCH codes, Goppa codes, and LDPC codes.

2.1 Network Coding

Consider a communication system with possibly many senders and receivers, interconnected by nodes, called a *network*. The senders inject *packets* of information into the network, and the packets travel throughout the network via the node connections. The behavior of the nodes largely determines the dynamics of the network. Figure 2.2 shows the general setup for the network model.

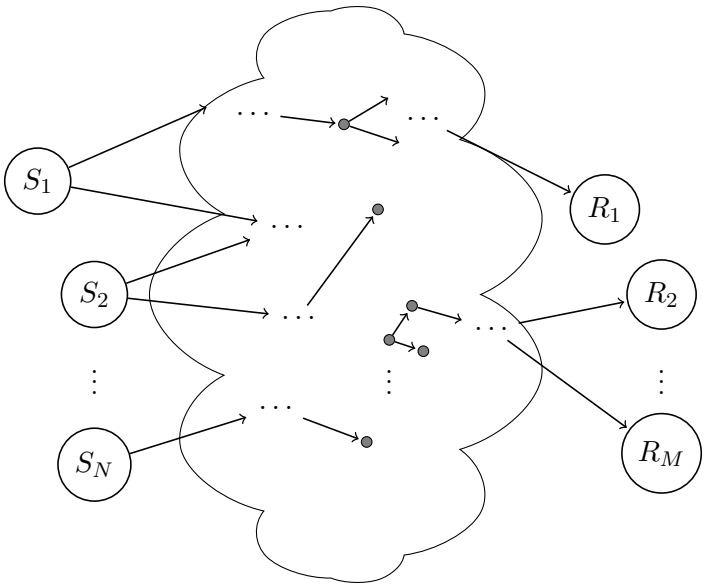

Figure 2.2: Network Model

Consider the case when nodes are simply allowed to route information and perform no algebraic operations. Then, if a node receives multiple packets of data, it must choose one to send forward. The canonical example of why routing is not an optimal solution is the butterfly network, given in Figure 2.3. In this network, the source, S, wishes to communicate the messages $\boldsymbol{a}$ and $\boldsymbol{b}$ to the receivers. If the source sends $\boldsymbol{a}$ to the first node and b to the second, then the third node receives both $\boldsymbol{a}$ and $\boldsymbol{b}$ and must make a choice. If node 3 chooses $\boldsymbol{x} = \boldsymbol{a}$, then the first receiver recovers only $\boldsymbol{a}$, while the second recovers $\boldsymbol{a}$ and $\boldsymbol{b}$. Similarly, if $\boldsymbol{x} = \boldsymbol{b}$ then the first recover obtain both $\boldsymbol{a}$ and $\boldsymbol{b}$ and the second receiver only $\boldsymbol{b}$. In either case, one receiver does not receive all the information.

Now, suppose that the nodes can take linear combinations of incoming packets. In this case, if node 3 receives both $\boldsymbol{a}$ and $\boldsymbol{b}$ and forwards $\boldsymbol{a} + \boldsymbol{b}$, then the both receivers can deduce $\boldsymbol{a}$ and $\boldsymbol{b}$ from linear combinations of the received packets. It was shown in [2] that in the case of a single source, the maximum possible throughput of a network can be achieved by allowing nodes to perform algebraic operations with a large enough alphabet size.

The process of allowing nodes to perform algebraic operations is often called *network coding*. In the case when the nodes perform only linear operations, it is often called *linear network coding*. Suppose that the topology of the network is fixed and known, i.e. the connections between nodes do not change with time and the senders know the directed graph formed by the nodes. This is sometimes called *coherent* network coding. For simplicity, consider only the case of one sender and one receiver. In this case and in the absence of errors, the networks can be described by a matrix relation. Suppose that S is the sender and R the receiver. S sends packets $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M \in \mathbb{F}_q^n$ and R will receive

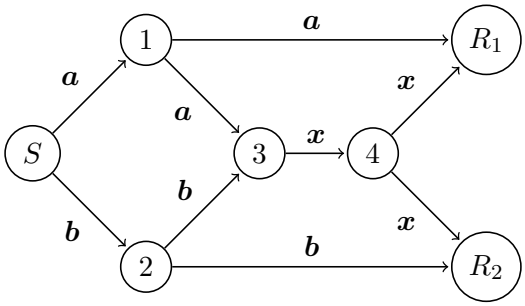

Figure 2.3: Butterfly Network

packets $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N \in \mathbb{F}_q^n$ according to the relation

$$H \underbrace{\begin{pmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_M \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} \boldsymbol{b}_1 \\ \vdots \\ \boldsymbol{b}_N \end{pmatrix}}_{B},$$

for some $H \in \mathbb{F}_q^{N \times M}$.

In the process of forwarding these combinations, transmission errors may occur between the nodes. We will assume that the transmission errors are additive, as in the q-ary symmetric channel. The directed graph defined by the node connections can therefore be viewed as a channel, however, an error occurring between two nodes in a network has the possibility of propagating and affecting all or many packets of information. Therefore, the Hamming distance would be inappropriate for modeling the errors induced by this channel. More specifically, if the nodes occasionally make additive errors as in the q-ary symmetric channel, then R receives packets of the form $HA + Z$, where Z captures the information about the errors that occured in transmission. Since any error occurring between two nodes in the network continues to be passed along by any other node receiving a corrupted packet, linear combinations of the errors will eventually reach the receiver and the number of errors at the receiver will be bounded by the rank of Z. This motivates the following definition.

Definition 2.1.1. Let $A, B \in \mathbb{F}_q^{M \times N}$. The rank metric on $\mathbb{F}_q^{M \times N}$ is given by

$$d_{\mathrm{R}}(A, B) = \mathrm{rk}(A - B).$$

This indeed defines a metric, and therefore one can consider coding as a packing problem in this metric. Codes designed with respect to this metric are often called *Delsarte codes* or *rank metric codes*. We will discuss this metric more in Chapter 3.

We note that even in the absence of errors, if $M = N$ and the $\boldsymbol{b}_i$ are simply a permutation of the $\boldsymbol{a}_i$, the matrix A and B may have large rank distance. Kötter and Kschichang noticed in [44] that in general, the subspace spanned by the $\boldsymbol{a}_i$ is invariant with respect to a more general network model. This observation led them to suggest that

subspaces, rather than matrices should be the central objects for use in error correcting codes for networks. In other words, the codewords should be elements of $\mathcal{P}(n, \mathbb{F}_q)$, the *projective geometry* of $\mathbb{F}_q^n$, i.e. the set of all possible subspaces of $\mathbb{F}_q^n$. $\mathcal{P}(n, \mathbb{F}_q)$ can be made into a metric space in the following way.

Definition 2.1.2. Let $U, V \in \mathcal{P}(n, \mathbb{F}_q)$. The subspace distance is given by

$$d_{\mathrm{S}}(U, V) = \dim(U + V) - \dim(U \cap V).$$

Considering codes with respect to d_{S} relates problems of error correction in networks to many problems in finite geometry. For more information as well as a brief backgroud on open problems in this direction, the reader is referred to [23]. Often, codes defined with respect to the subspace metric are referred to as *network codes*, although we will refer to them as *subspace codes* to avoid ambiguity with other definitions.

Of particular interest in designing subspace codes is to consider only those codes, $\mathcal{C} \subset \mathcal{P}(n, \mathbb{F}_q)$, with all codewords of $\mathcal{C}$ having the same dimension. We will denote the (k, n)-*Grassmann* over $\mathbb{F}_q$, i.e. the set of subspaces of dimension k in $\mathbb{F}_q^n$, by $\mathrm{Gr}(k, \mathbb{F}_q^n)$. We can therefore consider subspace codes which satisfy $\mathcal{C} \subset \mathrm{Gr}(k, \mathbb{F}_q^n) \subset \mathcal{P}(n, \mathbb{F}_q)$ for some k. Such codes are often called *constant-dimension* subspace codes. If k is the dimension of each subspace in a constant-dimension subspace code, then the maximum possible minimum distance is $2k$. In [44], a Singleton-like bound was given for any constant-dimension subspace code, $\mathcal{C} \in \mathrm{Gr}(k, \mathbb{F}_q^n)$, having minimum distance d, as

$$|\mathcal{C}| \leq \begin{bmatrix} n - (d-2)/2 \\ \max(k, n-k) \end{bmatrix}_q, \tag{2.1}$$

where the q-binomial coefficient is as in Definition 3.1.5. An important family of constant dimension codes, the *spread codes* are optimal packings with respect to d_{S}. They arise in finite geometry from the study of *spreads*.

Definition 2.1.3. A collection, $\mathcal{S}$, of subspaces of $\mathrm{Gr}(k, \mathbb{F}_q^n)$ is called an $(n, k)_q$-spread if $U \cap V = \{0\}$ for every $U \neq V \in \mathcal{S}$ and

$$\bigcup_{U \in \mathcal{S}} U = \mathbb{F}_q^n.$$

It is well-known that spreads exist if and only if k divides n. In this case, the number of elements of the spread is

$$|S| = \frac{q^n - 1}{q^k - 1}.$$

One can see that spread codes attain the bound in (2.1). For more information on spreads, spread-like constructions, and decoding algorithms, the reader is directed to [52, 36]. It was noted, for instance, that spread codes are closely related to rank metric codes and can in fact be constructed from them. More generally, constant-dimension subspace codes are closely related to rank metric codes [32].

2.2 Public-Key Cryptography

Before the ubiquity of modern private and public key cryptography, secret messages were hidden using steganographic methods and simple ciphers. Historical examples abound of secret messages being concealed with invisible ink, tattoos, or embedded into physical objects. As a modern method, secret messages can be embedded into digital files by using the smallest bits of the data. Private key ciphers became more popular as the use of machines grew, since it became more feasible to encrypt and decrypt large messages. Perhaps the most famous example, the Enigma machine, was used to encrypt Nazi communication during WWII until it was broken by allied cryptanalysts. Modern versions include algorithms such as the Advanced Encryption Standard (AES). Private key ciphers share the same drawback–the key must first be known to both the sender and receiver. Ideally, two parties would share the key through the use of a secure channel, however secure channels are difficult or impossible to establish in practice. Ultimately, the interested parties must communicate the private key in the presence of an adversary, hence, the need to communicate secretly in public has led to great interest in public key cryptosystems. These algorithms are generally much slower to encrypt and decrypt messages than in private key systems. Therefore, a combination of the two is often used; a public key cryptosystem is used to send a private key, which is used in subsequent communication between the two parties.

The asymmetric cryptography model, give in Figure 2.4 ignores the possibility of channel errors and focuses instead on the issue of secure communication–communication in the presence of an eavesdropper. In this model, the receiver devises an encryption rule which is communicated to the sender via the public key. The receiver does not publish the private information that facilitates the decryption. Without the private information, it should be extremely difficult for an eavesdropper to decrypt the ciphertexts. All asymmetric cryptography schemes therefore rely on the same principle–finding a problem in which some private information allows the intended recipient to efficiently solve, but without that private information the problem is infeasible.

The study of the difficulty of solving certain problems belongs to the field of complexity theory, and is one of fundamental importance to public key cryptography. Problems which are deemed "efficiently solvable" would be bad candidates for basing a public key cryptosystem upon, since they would provide little security against an eavesdropper. In this section, we will give an introduction to some basic notions in complexity theory which are the foundation for a heuristic about the difficult of solving certain problems. A complete survey of results in this direction would be a tedious digression, but a short introduction can be valuable to understand the context of public key cryptography. All statements and further elaboration can be found in [75].

Problems will be separated into two types. A *decision* problem is one in which the answer is given by 'yes' or 'no'. For instance, 'is n a prime number?', or 'is the graph G 3-colorable?' are both decision problems. A *search problem* is one in which the answer can not be given by a 'yes' or 'no'. For instance, 'What is the factorization of n?' is a search problem. Decision and search problems are often related. For instance, one can

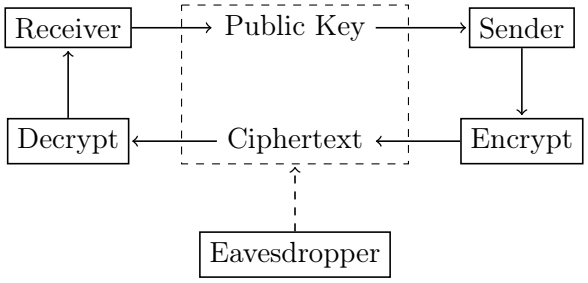

Figure 2.4: Asymmetric Cryptography Model

solve the problem 'what is the factorization of n?' if one can also solve 'is r a factor of n' for any r.

Definition 2.2.1. A problem, P, is called *polynomial*, if there exists a polynomial, f, such that for any instance $p \in P$, the number of operations required to solve p is bounded by $f(n)$, where n is the size of the input. We say that P is *polynomial in* n.

The class of all problems which are polynomial is denoted by **P**. If every problem in P with input size n can be solved in $f(n)$ operations, then we say that $P \sim O(f)$. We should note that the operations allowed are only those which can be performed on a deterministic Turing machine. We will omit the descriptor, deterministic. For practical purposes, we are interested in algorithms which are also allowed to use some element of randomness in their algorithms. Therefore, we will also say that a randomized algorithm is polynomial time if, on average, a solution can be found in polynomial time. Problems which can be solved in polynomial time by a randomized algorithm do not necessarily belong to **P**.

Some examples of problems which are in **P** include ordering a set of n integers, computing the inverse of a matrix over a finite field, and testing an integer for primality [1]. Not all problems are known to be in **P**. For instance, the best known algorithm for factoring an integer of size n bits has complexity is $O(2^{n^{1/3}}n^{2/3})$ [47]. From the intuition that exponential functions grow significantly faster than polynomial ones (at least asymptotically), in a cryptography context, one would want to design systems in which the best known algorithms for breaking the system require a number of steps exponential in the size of the input. This gives us the following definition:

Definition 2.2.2. An algorithm for solving a problem, P, will be said to have *exponential running time*, if there exists $\varepsilon > 0$ such that

$$P \sim O(2^{\varepsilon n}),$$

for an input of size n.

Again, we include randomized algorithms in this definition.

Definition 2.2.3. A decision problem, P, is said to be in **NP** if there exists a non-deterministic polynomial time algorithm capable of solving P.

Often, an equivalent definition is given in terms of a polynomial time certificate. A polynomial time 'yes' certificate is some information that allows one–if the solution is 'yes'–to verify in polynomial time that the solution is indeed 'yes'. In these terms, a decision problem, P is said to be in **NP** if there exists a polynomial time 'yes' certificate. In other words, if you are told that the answer to the decision problem is 'yes', then a 'yes' certificate would allow you to prove that the answer is indeed 'yes' in polynomial time. **NP** is then the class of decision problems in which a solution can be verified in polynomial time.

Example 2.2.4. Consider the following problem: 'For any given linear code, $\mathcal{C} \subset \mathbb{F}_q^n$, any point $\boldsymbol{y} \in \mathbb{F}_q^n$, and any integer $0 < t < n$, does there exist $\boldsymbol{x} \in \mathcal{C}$ such that $d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y}) = t$?' A 'yes' certificate in this case could be a point, $\boldsymbol{x}$ satisfying $d_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{y}) = t$. The verifier can check in polynomial time that $\boldsymbol{x}$ is indeed in $\mathcal{C}$, and that the distance is as required. Therefore, this problem is **NP**.

Clearly, $\mathbf{P} \subseteq \mathbf{NP}$, as any algorithm itself acts as a 'yes' certificate. An important subclass of the **NP** problems are the so called **NP**-*complete* problems. A decision problem $\mathcal{P}$ is **NP**-complete if any other decision problem in **NP** can be reduced to $\mathcal{P}$ in polynomial time. That is, a decision for $\mathcal{P}$ can be used to efficiently obtain a decision for any other problem in **NP**. Intuitively, these are the "hardest" **NP** problems.

It is widely believed that $\mathbf{P} \neq \mathbf{NP}$. If this is true, then any problem which is **NP**-complete is not solvable by a deterministic polynomial-time algorithm. Otherwise, every other problem in **NP** would be reducible to it, and hence $\mathbf{P} = \mathbf{NP}$. A class of problems which are more difficult than the **NP**-complete problem are the **NP**-*hard* problems. These are the problems (not necessarily in **NP**) which are intuitively *at least* as hard as the **NP**-complete problems.

Since it is commonly believed that **NP**-complete problems are very difficult–in that it is believed that there is no polynomial time algorithm solving any problem which is **NP**-complete–one can design cryptosystems attempting to exploit this.

One of the most thoroughly investigated asymmetric cryptography schemes is that of the RSA cryptosystem. The problem upon which the security of RSA is based is the problem of factoring integers which are the product of two prime numbers (this is a sub-problem of the general factorization problem). If $n = pq$ is the prime factorization, then a 'yes' certificate for the decision problem 'is n the product of two prime numbers?' can be, for instance, p or q. It is then easy to verify the 'yes' solution, so this decision problem is in **NP**. It is not known to be **NP**-complete and is in fact believed not to be. Because of its ubiquity and relative simplicity, we will present the RSA cryptosystem as an example [66].

Example 2.2.5 (Rivest-Shamir-Adelman)**.** Suppose that Bob wants to send a message to Alice. Alice chooses two large prime numbers, p and q and computes $n = pq$. Then, Alice finds an integer e, relatively prime to $\varphi(n) = (p-1)(q-1)$, the Euler φ function. That is, e is invertible in the ring $\mathbb{Z}/(pq\mathbb{Z})$. Let $d \equiv e^{-1} \mod \varphi(n)$. Alice makes public the key,

$$\kappa_{\mathrm{pub}} = (n, e),$$

and withholds the private key,

$$\kappa_{\text{pvt}} = (p, q, d).$$

Bob then chooses an element $m \in \mathbb{Z}/(n\mathbb{Z})$ to represent his message, and sends

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

to Alice through the public channel. Alice computes

$$c^d = m^{ed} \equiv m \mod n,$$

since $ed \equiv 1 \mod \varphi(n)$. If the eavesdropper can factor n and obtain p and q, then the value d can be computed from p, q, and e and therefore the cryptosystem is broken.

In Section 5.1, we will consider a subproblem of the MinRank problem which is known to be **NP**-complete problem. This subproblem, called the rank syndrome decoding problem, is conjectured to also be difficult although it is unknown whether it is **NP**-complete or not. It was first proposed for use in cryptography in [28].

2.3 Fuzzy Cryptosystems

In many public and private key cryptosystems, only an exact knowledge of the key allows one to gain access. There are conceivable instances in which this is not desirable. Consider, for instance, the following scenario. Bob is about to die, and he wants to bequeath his belongings only to those who, in life, were his true friends. In order to determine if a friend is a true friend or a false friend, they must correctly answer some questions regarding Bob's preferences. For instance, Bob could give Alice the query, 'what are my favorite movies?'. Bob cannot expect Alice to guess all of them correctly, and indeed Bob cannot be expected to list all his favorite movies without accidental omission. Nevertheless, he wants Alice to gain access if she can guess enough movies without guessing too many incorrect ones (for instance, she should not be given access if she lists all movies in existence). Bob will be dead when this process occurs, though, so he cannot tell Alice if she is correct or not; there must be an automated way to determine if she deserves the inheritance. Also, Bob can not write down the list of his favorite movies in an easily distinguishable form or else it may be stolen.

The idea is to devise a system so that the system releases the correct key if and only if Alice's answer is close to Bob's. Any party attempting to access the system will be called a *witness*. An unintended party will be called an *adversary* and Alice (or any intended party) will be called an *authentic* user. Bob's data will be called the *template*. The data itself will be composed of *features*. We say that a witness *decommits* the key if the witness is a valid set of features for unlocking the system.

Another–more feasible and less vindictive–scenario arises in the problem of biometric authentication. Biometrics are used to refer to any property of the body which is relatively invariant with respect to time and can be used to uniquely identify an individual

with high probability. Of course, the usual suspects–fingerprint and iris data–are biometrics, but even processes like handwriting or speech are considered biometrics as well. In these cases, even an authentic user often cannot perfectly reproduce the exact same features in a reliable way.

An important precusor to the fuzzy vault scheme considered in Chapter 6 is the fuzzy commitment scheme of Juels and Wattenberg [41]. Let $\mathcal{C} \in \mathbb{F}_q^n$ be a t-error correcting code with respect to the Hamming metric with an efficient decoding algorithm, $\mathtt{dec}\colon \mathbb{F}_q^n \to \mathbb{F}_q^n$, and let $h\colon \mathbb{F}_q^n \mapsto \mathbb{F}_q^\ell$ be a hash function. Let $\boldsymbol{x} \in \mathbb{F}_q^n$ be the template and let $\boldsymbol{\delta}$ be any vector such that $\boldsymbol{c} = \boldsymbol{x} + \boldsymbol{\delta} \in \mathcal{C}$. The fuzzy commitment, F, is then defined by

$$F(\boldsymbol{x}, \boldsymbol{\delta}) = (h(\boldsymbol{c}), \boldsymbol{\delta}).$$

In this case, the codeword $\boldsymbol{c}$ is the key.

Lemma 2.3.1. *Let $\boldsymbol{x}'$ be such that $d_{\mathrm{H}}(\boldsymbol{x}', \boldsymbol{x}) \leq t$. Then for any $\boldsymbol{c} \in \mathcal{C}$, the witness $\boldsymbol{x}'$ can be used to decommit $F(\boldsymbol{c}, \boldsymbol{x})$ successfully.*

Proof. If $d_{\mathrm{H}}(\boldsymbol{x}', \boldsymbol{x}) \leq t$, then $d_{\mathrm{H}}(\boldsymbol{x}' + \boldsymbol{\delta}, \boldsymbol{x} + \boldsymbol{\delta}) \leq t$. Since $\boldsymbol{x} + \boldsymbol{\delta} = \boldsymbol{c}$ is a codeword of $\mathcal{C}$, we obtain $\mathtt{dec}(\boldsymbol{x}' + \boldsymbol{\delta}) = \boldsymbol{c}$. Simply checking the value of $h(\boldsymbol{c})$ verifies that $\boldsymbol{x}'$ was a valid witness. □

If an adversary submits $\boldsymbol{y}$ without knowledge of the template, then we do not expect $d_{\mathrm{H}}(\boldsymbol{y} + \boldsymbol{\delta})$ to be within distance t of $\boldsymbol{x} + \boldsymbol{\delta}$. Therefore, decoding will not yield $\boldsymbol{c}$ and so $h(\mathtt{dec}(\boldsymbol{y} + \boldsymbol{\delta})) \neq h(\boldsymbol{c})$. We then reject this witness.

One principal disadvantage to the fuzzy commitment scheme is that the decommission of the key is not permutation invariant with respect to the features. In the scenario of Bob's bequeathment, the order of Alice's answers should not affect the decommission process. In other words, the fuzzy commitment scheme works well when the errors occuring during feature extraction are amenable to the Hamming distance, but can fail quite easily, otherwise. Other examples of primitives for fuzzy cryptosystems include the fuzzy syndrome hashing scheme [51] and fuzzy extractors [22], and the fuzzy vault [42] which is examined in Chapter 6.

Chapter 3

Rank Metric Codes

Rank metric codes were introduced by Delsarte in [18] using association schemes and proposed more recently under a different framework by Gabidulin [26]. Delsarte codes are generalizations of the latter; their relationship and duality theory are described in [65]. In this paper, we will consider rank metric codes defined in an extension field, $\mathbb{E}$, over a base field $\mathbb{F}$. The degree of the extension will always be denoted by $[\mathbb{E}\colon \mathbb{F}] = m$. We only consider finite fields, i.e. $\mathbb{F} = \mathbb{F}_q$ and $\mathbb{E} = \mathbb{F}_{q^m}$ for some prime power, q.

Of particular interest are codes meeting the Singleton bound in the rank metric, called maximum rank distance (MRD) codes. The first examples of such codes were given in [18]. In [26], a family of such codes were constructed using linearized polynomials which was later generalized in [45]. Sheekey in [70] constructed a new class of codes closely related to Gabidulin codes which was later generalized in [50]. The codes arising from the generalized Sheekey construction are in general not linear over the extension field, but contain as special cases all the known linear (over the extension field) MRD codes. Aside from these,there are no other known codes satisfying the Singleton bound which are linear over the extension field $\mathbb{F}_{q^m}$. Other constructions exist which are non-linear or linear only over a subfield of $\mathbb{F}_{q^m}$.

In this chapter, we introduce the rank metric and some fundamental tools that we will use throughout the paper. In Section 3.1, we summarize some of the fundamental results regarding rank-metric codes with an emphasis on the relationship between rank metric codes and the coordinate-wise Frobenius map. We also give the characterization of the linear, semi-linear, and $\mathbb{F}_q$-linear isometries due to [57]. In Section 3.2, we proceed to give some examples of rank metric codes. First, we present an important family of codes for rank based public key cryptography–generalized Gabidulin codes. We also present the construction of Sheekey. We then give some background for our construction of new MRD codes. Our construction is sporadic and constructed by combinatorial means, nevertheless, it proves a new construction of linear MRD codes which is not semi-linearly equivalent to a generalized Gabidulin code. The results in this direction can be found in [39]. Lastly, we present the low rank parity check (LRPC) codes, which are proposed in [31] as a feasible alternative to Gabidulin based cryptosystems.

3.1 Preliminaries

Let $\mathbb{F}_{q^m}$ be an extension field over $\mathbb{F}_q$. For a fixed basis, say $\{b_1, ..., b_m\} \subset \mathbb{F}_{q^m}$ of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$, any element $a \in \mathbb{F}_{q^m}$ can be written uniquely as

$$a = \sum_{i=1}^{m} a_i b_i,$$

for $a_i \in \mathbb{F}_q$. Hence, we can represent a as the column vector $[a_1, \ldots, a_m]^T \in \mathbb{F}_q^{m\times 1}$ in the canonical way. Given an element $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$, we can expand each coordinate into a column according to this fixed basis, obtaining a matrix $[\boldsymbol{x}] \in \mathbb{F}_q^{m\times n}$. In this way, any subset $\mathcal{C} \subset \mathbb{F}_{q^m}^n$ can be viewed as a subset of $\mathbb{F}_q^{m\times n}$. These are sometimes called *Delsarte (matrix) codes* [65]. We will consider mostly the case when $\mathcal{C} \subset \mathbb{F}_{q^m}^n$ is $\mathbb{F}_{q^m}$-linear and we will call these *linear rank metric* codes as they agree with the alphabet of the code when codewords are viewed as vectors. In order to distinguish those codes which are linear over a subfield $\mathbb{F}' \subsetneq \mathbb{F}_{q^m}$, we will specify that the code is $\mathbb{F}'$-linear. Clearly, all linear rank metric codes are $\mathbb{F}_q$-linear Delsarte codes, but not all $\mathbb{F}_q$-linear Delsarte codes are linear rank metric codes.

We can therefore define the rank metric for vectors to coincide with Definition 2.1.1 for matrices.

Definition 3.1.1. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}_{q^m}^n$. The *rank distance* between $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by

$$d_{\mathrm{R}}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{rk}\left([\boldsymbol{x} - \boldsymbol{y}]\right). \tag{3.1}$$

Lemma 3.1.2. *The map $d_{\mathrm{R}} \colon \mathbb{F}_{q^m}^n \times \mathbb{F}_{q^m}^n \mapsto \{0, \ldots, m\}$ given in (3.1) defines a metric. Moreover, the distance is independent of the choice of basis for $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$.*

Proof. First, fix a basis $\mathcal{B} = \{b_1, ..., b_m\}$ for $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ and let $\boldsymbol{x}, \boldsymbol{y}$, and $\boldsymbol{z} \in \mathbb{F}_{q^m}^n$. It is clear that

$$d_{\mathrm{R}}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{rk}([\boldsymbol{x} - \boldsymbol{y}]) = \mathrm{rk}([\boldsymbol{y} - \boldsymbol{x}]) = d_{\mathrm{R}}(\boldsymbol{y}, \boldsymbol{x}).$$

Also,

$$d_{\mathrm{R}}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{rk}([\boldsymbol{x} - \boldsymbol{y}]) \geq \mathrm{rk}([\boldsymbol{0}]) = 0,$$

with equality if and only if $\boldsymbol{x} = \boldsymbol{y}$. For any matrices A, B of the same dimensions, $\mathrm{rk}(A + B) \leq \mathrm{rk}(A) + \mathrm{rk}(B)$. Therefore, we have

$$\begin{aligned} d_{\mathrm{R}}(\boldsymbol{x}, \boldsymbol{y}) &= \mathrm{rk}([\boldsymbol{x} - \boldsymbol{y}]) \\ &= \mathrm{rk}([\boldsymbol{x} - \boldsymbol{z}] + [\boldsymbol{z} - \boldsymbol{y}]) \\ &\leq \mathrm{rk}([\boldsymbol{x} - \boldsymbol{z}]) + \mathrm{rk}([\boldsymbol{z} - \boldsymbol{y}]) \\ &= d_{\mathrm{R}}(\boldsymbol{x}, \boldsymbol{z}) + d_{\mathrm{R}}(\boldsymbol{z}, \boldsymbol{y}). \end{aligned}$$

Therefore, d_{R} defines a metric. To show the value is independent of the choice of basis, we note that for any change of basis matrix $S \in \mathrm{GL}_m(\mathbb{F}_q)$, we have $\mathrm{rk}(S[\boldsymbol{x}]) = \mathrm{rk}([\boldsymbol{x}])$. □

The *rank weight*, or *rank* of a vector $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ is simply given by

$$\mathrm{wt}_{\mathrm{R}}(\boldsymbol{x}) = \mathrm{rk}(\boldsymbol{x}) = d_{\mathrm{R}}(\boldsymbol{x}, \boldsymbol{0}).$$

Other ways to conceive of the rank of a vector $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ is as the dimension of the $\mathbb{F}_q$-vector space spanned by the coordinates of $\boldsymbol{x}$, or the number of coordinates of $\boldsymbol{x}$ which are linearly independent over $\mathbb{F}_q$. These can all be seen to be equivalent.

There is the possibility for ambiguity in the different notions of rank. However, it should be clear by context what is meant. If $M \in \mathbb{F}_{q^m}^{m \times n}$ is a matrix, by the rank of M, we mean the rank in the usual way; the number of linearly independent rows or columns of M over $\mathbb{F}_{q^m}$. It is known that the row rank and the column rank of M are equal when considered over $\mathbb{F}_{q^m}$. If M is a matrix with entries in $\mathbb{F}_{q^m}$, we will also consider the $\mathbb{F}_q$-span of the columns of M, the dimension of which will be called the column rank (over $\mathbb{F}_q$). It is not true that the column rank of M over $\mathbb{F}_q$ will equal the row rank over $\mathbb{F}_q$ if M has entries in $\mathbb{F}_{q^m}$. We will denote the column rank of M over $\mathbb{F}_q$ by $\mathrm{colrk}_{\mathbb{F}_q}(M)$. If $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$, by the rank of $\boldsymbol{x}$, we mean the rank weight of $\boldsymbol{x}$.

Denote the rank sphere of radius r around $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ by

$$S_{r,n}^{\mathrm{R}}(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{F}_{q^m}^n \mid d_{\mathrm{R}}(\boldsymbol{x}, \boldsymbol{y}) = r\}, \tag{3.2}$$

and the rank ball of radius r around $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ by

$$B_{r,n}^{\mathrm{R}}(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{F}_{q^m}^n \mid d_{\mathrm{R}}(\boldsymbol{x}, \boldsymbol{y}) \leq r\} = \bigcup_{i=0}^{r} S_{r,n}^{\mathrm{R}}(\boldsymbol{x}). \tag{3.3}$$

From Definition 3.1.1 we can observe that for any $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$, $\mathrm{wt}_{\mathrm{R}}(\boldsymbol{x}) \leq \mathrm{wt}_{\mathrm{H}}(\boldsymbol{x})$, and therefore it follows that for any $0 \leq r \leq \min\{m, n\}$,

$$B_{r,n}^{\mathrm{H}}(\boldsymbol{x}) \subseteq B_{r,n}^{\mathrm{R}}(\boldsymbol{x}).$$

In a restricted way, the rank metric generalizes the Hamming metric; any subset of $\mathbb{F}_{q^m}^n$ viewed as a code in the rank metric capable of correcting t errors will also correct t errors in the Hamming metric. On the other hand, the same subset may be able to correct more errors when viewed directly as a code in the Hamming metric.

In order to determine the cardinality of $B_{r,n}^{\mathrm{R}}(\boldsymbol{0}) \subset \mathbb{F}_{q^m}^n$, we use the following lemma.

Lemma 3.1.3. *Suppose $r \leq n$. Then, the map*

$$\varphi \colon S_{r,r}^{\mathrm{R}}(\boldsymbol{0}) \times M_{r \times n}(\mathbb{F}_q)/\mathrm{GL}_r(\mathbb{F}_q) \longrightarrow S_{n,r}^{\mathrm{R}}(\boldsymbol{0}),$$

$$(\boldsymbol{v}, U) \longmapsto \boldsymbol{v}U$$

is a bijection.

Proof. As representatives of the cosets in $M_{r \times n}(\mathbb{F}_q)/\mathrm{GL}_r(\mathbb{F}_q)$ we consider the reduced row echelon form of the respective row span of the elements of the coset.

We first show that φ is surjective. For this consider an arbitrary element in the image of φ, i.e. a vector $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ of rank r, and let $x_{i_1}, \ldots, x_{i_r}$ be the first r independent entries of $\boldsymbol{x}$, in positions $i_1, \ldots, i_r$. Then, the remaining $n-r$ entries of $\boldsymbol{x}$ can be expressed as an $\mathbb{F}_q$-linear combination of $x_{i_1}, \ldots, x_{i_r}$, thus we can write $\boldsymbol{x} = (x_{i_1}, \ldots, x_{i_r})M$ for some matrix $M \in \mathbb{F}_q^{t \times n}$. Then there exists $S \in \mathrm{GL}_r(\mathbb{F}_q)$ such that $U = S^{-1}M$ is in reduced row echelon form. We get $(x_{i_1}, \ldots, x_{i_r})S \in S_{r,r}^{\mathrm{R}}(\mathbf{0})$ and $\boldsymbol{x} = \varphi((x_{i_1}, \ldots, x_{i_r})S, U)$, thus φ is surjective.

To show injectivity, suppose that there are two pre-images, i.e. $\boldsymbol{x} = \varphi(\boldsymbol{v}, U) = \varphi(\boldsymbol{v'}, U')$. Without loss of generality, we can assume that $U = [I_r \mid *]$. Denote by U'_j the jth column of U'. Then we have

$$(x_1, \ldots, x_r) = \boldsymbol{v} = (\boldsymbol{v'}U'_1, \ldots, \boldsymbol{v'}U'_r).$$

Since $\boldsymbol{v}$ has rank r, $U'_1, \ldots, U'_r$ must be non-zero. Because U' is in reduced row echelon form, we get $U' = [I_r \mid *]$ and hence

$$(x_1, \ldots, x_r) = \boldsymbol{v} = \boldsymbol{v'}.$$

We furthermore have $x_j = \boldsymbol{v}U_j = \boldsymbol{v'}U'_j$ for $j = r+1, \ldots, n$. Thus

$$\boldsymbol{v}U_j = \boldsymbol{v'}U'_j \Leftrightarrow \boldsymbol{v}U_j = \boldsymbol{v}U'_j \Leftrightarrow \boldsymbol{v}(U_j - U'_j) = 0.$$

Since $\mathrm{rk}(\boldsymbol{v}) = r$, we get $U_j - U'_j = 0$ for $j = r+1, \ldots, n$. Thus $U = U'$ and we have shown that φ is injective. □

Lemma 3.1.3 was observed, for instance, in [33, 73]. The rank metric can also be studied over other fields than finite fields [67], although we only consider the finite field case.

Definition 3.1.4. Let V be a vector space of finite dimension, n, over some field $\mathbb{K}$. The (k,n)-Grassmannian of V, $\mathrm{Gr}(k, V)$ is the space of k-dimensional $\mathbb{K}$-subspaces of V.

For $k \leq n$, each element of $\mathrm{Gr}(k, \mathbb{F}_q^n)$ can be represented as the row span of some full rank matrix belonging to $\mathbb{F}_q^{k \times n}$. This representation can be made unique by, for instance, considering the reduced row echelon form matrix associated to each subspace $U \in \mathrm{Gr}(k, \mathbb{F}_q^n)$.

Definition 3.1.5. The q-binomial coefficient, or Gaussian binomial coefficient, is the number of r-dimensional $\mathbb{F}_q$-subspaces of $\mathbb{F}_q^n$. For $r \leq n$, it is given by

$$\begin{bmatrix} n \\ r \end{bmatrix}_q = \prod_{i=0}^{r-1} \frac{q^n - q^i}{q^r - q^i}.$$

Corollary 3.1.6. *We have,*

$$\left|B_{r,n}^{\mathrm{R}}(\mathbf{0})\right| = \sum_{i=0}^{r} \begin{bmatrix} n \\ i \end{bmatrix}_q \prod_{j=0}^{r-1} (q^m - q^j) \tag{3.4}$$

Proof. We first observe that

$$|S^{\mathrm{R}}_{r,r,}(\mathbf{0})| = \prod_{j=0}^{r-1}(q^m - q^j).$$

Then, using Lemma 3.1.3, we have that

$$\begin{aligned}\left|B^{\mathrm{R}}_{r,n}(\mathbf{0})\right| &= \sum_{i=0}^{r} |S^{\mathrm{R}}_{r,n}(\mathbf{0})| \\ &= \sum_{i=0}^{r} |S^{\mathrm{R}}_{r,r}(\mathbf{0})| \cdot |M_{t\times n}(\mathbb{F}_q)/\mathrm{GL}_t(\mathbb{F}_q)| \\ &= \sum_{i=0}^{r} \begin{bmatrix} n \\ i \end{bmatrix}_q \prod_{j=0}^{r-1}(q^m - q^j).\end{aligned}$$

□

Corollary 3.1.6 can be found in [33] with the connection with the Grassmannian given in [76]. We will need to establish some preliminary notation that will be used throughout this paper. Let $M = (M_{a,b})$ be a matrix (or vector) over $\mathbb{F}_{q^m}$. If $\mathbb{K}$ is a sub-field or extension field of $\mathbb{F}_{q^m}$, we denote the $\mathbb{K}$-span of the rows of M by $\langle M\rangle_{\mathbb{K}}$. For any non-negative integer i, we will use the shorthand, $[i]$, to mean the ith Frobenius power, q^i. Lastly, we will denote the coordinate-wise Frobenius map by, $M^{([i])} = (M^{[i]}_{a,b})$, that is, each entry of M raised to the $[i]$th power. For a subset $\mathcal{S} \subset \mathbb{F}^n_{q^m}$, we define $\mathcal{S}^{([i])} = \{\boldsymbol{s}^{([i])} \mid \boldsymbol{s} \in \mathcal{S}\}$.

The following proposition summarizes some important properties of the coordinate-wise Frobenius map.

Proposition 3.1.7. *Let $M \in \mathbb{F}^{a\times b}_{q^m}$, $i \in \mathbb{Z}$.*

1. $\langle M\rangle^{([i])}_{\mathbb{F}_{q^m}} = \langle M^{([i])}\rangle_{\mathbb{F}_{q^m}}$,
2. *If $M \in \mathbb{F}^{a\times b}_{q} \subset \mathbb{F}^{a\times b}_{q^m}$, then $M = M^{([i])}$,*
3. *If $N \in \mathbb{F}^{b\times c}_{q^m}$, then $(MN)^{([i])} = M^{([i])}N^{([i])}$.*

Proof. To prove statement 1, we show that $\langle M\rangle^{([1])}_{\mathbb{F}_{q^m}} = \langle M^{([1])}\rangle_{\mathbb{F}_{q^m}}$ from which the result follows by repeated applications of the coordinate-wise Frobenius map. Let $M_1, \ldots, M_a$ be the rows of M. Any element $\boldsymbol{m} \in \langle M\rangle^{([i])}_{\mathbb{F}_{q^m}}$ can be written

$$\boldsymbol{m} = \left(\sum_{i=1}^{a} \alpha_i M_i\right)^{([1])} = \sum_{i=1}^{a} \alpha_i^{([1])} M_i^{([1])} \in \langle M^{([1])}\rangle_{\mathbb{F}_{q^m}}.$$

Similarly, for $\boldsymbol{m} \in \langle M^{([1])} \rangle_{\mathbb{F}_{q^m}}$,

$$\boldsymbol{m} = \sum_{i=1}^{a} \alpha_i M_i^{([1])} = \left(\sum_{i=1}^{a} \alpha_i^{([m-1])} M_i \right)^{([1])} \in \langle M \rangle_{\mathbb{F}_{q^m}}^{([i])},$$

since for any $\alpha \in \mathbb{F}_{q^m}$, $(\alpha^{([m-1])})^{([1])} = \alpha$. Statement 2 follows since the Frobenius automorphism, $\sigma \colon x \mapsto x^{[1]}$, fixes elements of $\mathbb{F}_q$, and statement 3 follows from σ being an automorphism of $\mathbb{F}_{q^m}$.

□

The following properties of the coordinate-wise Frobenius map will be used throughout the paper. The first statement follows straightforwardly from the $\mathbb{F}_q$-linearity of the Frobenius map, the second and the third follow immediately from Proposition 3.1.7 and can be found, for instance, in [35, 39].

Lemma 3.1.8. *The following hold for any prime power $[i] = q^i$ and $0 < n \leq m$.*

1. *Let $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ have rank r. Then, $\boldsymbol{x}^{([i])}$ also has rank r.*
2. *Let $M \in \mathrm{GL}_n(\mathbb{F}_{q^m})$. Then, $(M^{-1})^{([i])} = (M^{([i])})^{-1}$.*
3. *Let $\mathcal{S} \subset \mathbb{F}_{q^m}^n$ be an $\mathbb{F}_{q^m}$-subspace. Then, $\mathcal{S}^{([1])} = \mathcal{S}$ if and only if $\mathcal{S}$ has a basis contained in $\mathbb{F}_q^n$.*

Corollary 3.1.9. *Let $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ have rank r with decomposition $\boldsymbol{x} = \hat{\boldsymbol{x}} U$ according to Lemma 3.1.3. Then,*

$$\langle \boldsymbol{x}, \boldsymbol{x}^{([1])}, \ldots, \boldsymbol{x}^{([r-1])} \rangle_{\mathbb{F}_{q^m}} = \langle U \rangle_{\mathbb{F}_{q^m}}.$$

Proof. Since U has entries in $\mathbb{F}_q$, by Proposition 3.1.7, $\boldsymbol{x}^{([i])} = \hat{\boldsymbol{x}}^{([i])} U$ and therefore, $\boldsymbol{x}^{([i])} \in \langle U \rangle_{\mathbb{F}_{q^m}}$ for every i. We show that $\boldsymbol{x}, \ldots, \boldsymbol{x}^{([r-1])}$ are independent over $\mathbb{F}_{q^m}$. Suppose on the contrary that they are dependent. Then, since U has rank r, we would have

$$\boldsymbol{x}^{([r-1])} = \sum_{i=0}^{r-2} \gamma_i \boldsymbol{x}^{([i])} \in \langle \boldsymbol{x}, \ldots, \boldsymbol{x}^{([r-2])} \rangle_{\mathbb{F}_{q^m}} \subsetneq \langle U \rangle_{\mathbb{F}_{q^m}}.$$

We note that $\gamma_0 \neq 0$, otherwise $\boldsymbol{x}$ would have rank smaller than r, a contradiction. Then,

$$\langle \boldsymbol{x}, \ldots, \boldsymbol{x}^{([r-2])} \rangle_{\mathbb{F}_{q^m}}^{([1])} = \langle \boldsymbol{x}^{([1])}, \ldots, \boldsymbol{x}^{([r-1])} \rangle_{\mathbb{F}_{q^m}} = \langle \boldsymbol{x}, \ldots, \boldsymbol{x}^{([r-2])} \rangle_{\mathbb{F}_{q^m}}.$$

and by the third point of Lemma 3.1.8, $\langle \boldsymbol{x}, \ldots, \boldsymbol{x}^{([r-2])} \rangle_{\mathbb{F}_{q^m}}$ has a basis of elements in $\mathbb{F}_q$. However, this contradicts Lemma 3.1.3 since we could express $\boldsymbol{x} = \hat{\boldsymbol{x}} U'$, where $U' \in \mathbb{F}_q^{(r-1) \times n}$. □

Corollary 3.1.10. *Let $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ have rank r with decomposition $\boldsymbol{x} = \hat{\boldsymbol{x}} U$ according to Lemma 3.1.3. Then,*

$$\langle [\boldsymbol{x}] \rangle_{\mathbb{F}_{q^m}} = \langle U \rangle_{\mathbb{F}_{q^m}}.$$

Corollary 3.1.10 can be found in [43], where the space $\langle U \rangle_{\mathbb{F}_{q^m}}$ is called the rank support. From Corollary 3.1.9, we see that $r+1$ consecutive coordinate-wise Frobenius powers of $\boldsymbol{x}$ must be dependent, if $\boldsymbol{x}$ has rank r. Therefore, there exists some linear combination $\gamma_0, \ldots, \gamma_{r-1} \in \mathbb{F}_{q^m}$ such that

$$\boldsymbol{x}^{([r])} + \sum_{i=0}^{r-1} \gamma_i \boldsymbol{x}^{([i])} = 0, \tag{3.5}$$

from which we obtain the following definition.

Definition 3.1.11. Let $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ have rank r, and $\gamma_0, \ldots, \gamma_{r-1}$ be as in Equation (3.5). Then,

$$Z^{[r]} + \sum_{i=0}^{r-1} \gamma_i Z^{[i]},$$

is called the *annulator polynomial* of $\boldsymbol{x}$.

The annulator polynomial depends only on the $\mathbb{F}_q$-span of the coordinates of $\boldsymbol{x}$. In particular, if $\boldsymbol{x} = (x_1, \ldots, x_n)$, the unique smallest degree monic polynomial vanishing on the set $\langle x_1, \ldots, x_n \rangle_{\mathbb{F}_q} \subset \mathbb{F}_{q^m}$ is precisely the annulator polynomial. We note that $\dim(\langle x_1, \ldots, x_n \rangle_{\mathbb{F}_q}) = r$, so the smallest degree polynomial vanishing on $\langle x_1, \ldots, x_n \rangle_{\mathbb{F}_q}$ must have degree $[r]$. The annulator polynomial is defined this way, for instance, in [59, 30]. However, we choose our definition because it is constructive, rather than a statement of existence.

Because the annulator polynomial captures the information regarding the $\mathbb{F}_q$-subspace spanned by the coordinates of $\boldsymbol{x}$, we will call $\langle x_1, \ldots, x_n \rangle_{\mathbb{F}_q}$ the *vector space support* of $\boldsymbol{x}$. In [30], this is simply referred to as the support of $\boldsymbol{x}$, although we make the distinction for the following reason. Recall that if $\boldsymbol{x} = \hat{\boldsymbol{x}} U$, then U is determined up to a left action of $\mathrm{GL}_{\mathrm{rk}(U)}(\mathbb{F}_q)$. That is, $\langle U \rangle_{\mathbb{F}_q}$ is invariant with respect to the choice of decomposition of $\boldsymbol{x}$. Therefore, we can think of $\langle U \rangle_{\mathbb{F}_q}$ as encoding some information about the distribution, or location, of the values of $\boldsymbol{x}$. From Lemma 3.1.3, $\langle U \rangle_{\mathbb{F}_q}$ can be viewed as an element of the Grassmann space, $\mathrm{Gr}(r, \mathbb{F}_q^n)$, and therefore we will call $\langle U \rangle_{\mathbb{F}_q}$ the *Grassmann support* of $\boldsymbol{x}$. By abuse of notation, we will also call $\langle U \rangle_{\mathbb{F}_{q^m}}$ the Grassmann support of $\boldsymbol{x}$, since it is often more useful to consider this larger space. The analog between the Grassmann support in the rank metric and the coordinates of the error in the Hamming metric were observed to some degree in [61, 33].

To summarize, if $\boldsymbol{x} = \hat{\boldsymbol{x}} U$ as in Lemma 3.1.3, then $\langle x_1, \ldots, x_n \rangle_{\mathbb{F}_q} = \langle \hat{x}_1, \ldots, \hat{x}_r \rangle_{\mathbb{F}_q}$ is called the vector space support of $\boldsymbol{x}$ and $\langle U \rangle_{\mathbb{F}_q}$ is called the Grassmann support. These two notions almost uniquely determine $\boldsymbol{x}$. In order to have a bijection as in Lemma 3.1.3, we simply need to associate each space $\langle U \rangle_{\mathbb{F}_q}$ to a single matrix from $\mathrm{GL}_r(\mathbb{F}_q) U$. This is analogous to the case in the Hamming metric, where a vector is uniquely determined by the values of the coordinates, together with the location of those values. We will denote the vector space support and Grassmann support of $\boldsymbol{x}$ by $\mathrm{supp}_{\mathrm{vs}}(\boldsymbol{x})$ and $\mathrm{supp}_{\mathrm{Gr}}(\boldsymbol{x})$, respectively.

We can furthermore extend the definition of the Grassmann support for matrices. First, recall that for a matrix $M \in \mathbb{F}_{q^m}^{k\times n}$ of rank r, one can write

$$M = M'U,$$

where $M' \in \mathbb{F}_{q^m}^{k\times r}$ and $U \in \mathbb{F}_{q^m}^{r\times n}$ both having rank r. A slight modification of this idea gives the following:

Proposition 3.1.12. *Let $M \in \mathbb{F}_{q^m}^{k\times n}$ have column rank t over $\mathbb{F}_q$. Then, we can write*

$$M = M'U, \tag{3.6}$$

with $M' \in \mathbb{F}_{q^m}^{k\times t}$ and $U \in \mathbb{F}_q^{t\times n}$ of rank t. This decomposition is unique up to choice of representation of $\langle U\rangle_{\mathbb{F}_q}$.

Proof. Let $M_{i_1}, ..., M_{i_t}$ denote the first t independent columns of M. Then, every other column of M can be written as an $\mathbb{F}_q$-linear combination of these, so that for some $U \in \mathbb{F}_q^{t\times n}$,

$$M = \underbrace{\begin{bmatrix} M_{i_1} & M_{i_2} & \dots & M_{i_t} \end{bmatrix}}_{M'} U.$$

If $M = N'V$ is another such decomposition, then the $\mathbb{F}_q$-span of the columns of M' must equal the $\mathbb{F}_q$-span of the columns of N'. Therefore, there is an invertible matrix $S \in \mathrm{GL}_t(\mathbb{F}_q)$ such that $M'S = N'$. Then, we have

$$M'U = N'V = M'SV,$$

and therefore we must have $SV = U$, so they have the same row space. □

We will call $\langle U\rangle_{\mathbb{F}_{q^m}}$ the Grassmann support of M and denote it by $\mathrm{supp}_{\mathrm{Gr}}(M) = \langle U\rangle_{\mathbb{F}_{q^m}}$. To further justify extending the definition in this way, we observe the following:

Corollary 3.1.13. *Let $M \in \mathbb{F}_{q^m}^{k\times n}$ be such that $\mathrm{colrk}_{\mathbb{F}_q}(M) = t$ and let $\mathcal{U} \subset \mathbb{F}_{q^m}^n$ be the Grassmann support of M. Then, any element $\boldsymbol{x} \in \langle M\rangle_{\mathbb{F}_{q^m}}$ has Grassmann support contained in $\mathcal{U}$.*

One goal in coding theory is to classify codes by isometry. Let $\mathcal{M}$ be a vector space over a field $\mathbb{K}$ with distance function d^*. Recall that a map $f\colon \mathcal{M} \to \mathcal{M}$ is called an *isometry* if $d^*(\boldsymbol{m}) = d^*(f(\boldsymbol{m}))$, for each $\boldsymbol{m} \in \mathcal{M}$. The space of all isometries on M, $\mathrm{Iso}(\mathcal{M})$ is a group under composition. An important class of isometries are the the linear isometries, denoted by $\mathrm{Lin}(\mathcal{M}) \subset \mathrm{Iso}(\mathcal{M})$, which are the isometries which are also linear maps from $\mathcal{M}$ to $\mathcal{M}$. For a linear code $\mathcal{C}$ and any linear isometry, φ, $\varphi(\mathcal{C})$ is also a linear code and moreover if G is a generator matrix for $\mathcal{C}$, then $f(G)$ is a generator matrix for $\varphi(\mathcal{C})$. Hence we can define an equivalence relation on the set of linear codes by the classes arising from the action of these linear isometries. Codes $\mathcal{C}_1$ and $\mathcal{C}_2$ will be called *(linearly) equivalent*, if there exists a (linear) isometry, f, so that $f(\mathcal{C}_1) = \mathcal{C}_2$. Equivalent codes have identical parameters.

In the case of the Hamming metric in $\mathbb{F}_q^n$, the $\mathbb{F}_q$-linear isometries of $\mathbb{F}_q^n$ are known to be isomorphic to $\mathbb{F}_q^* \wr_n S_n$ [9]. This is realizable as multiplication on the right by diagonal matrices with all non-zero entries as well as right multiplication by a permutation matrix.

Morrison studied different isometry classes of rank metric codes in [57]. In particular, it was shown that

$$\mathrm{Lin}((\mathbb{F}_{q^m}^n, d_{\mathrm{R}})) \cong (\mathbb{F}_{q^m}^* \times \mathrm{GL}_n(\mathbb{F}_q))/N,$$

where $N = \{(\lambda, \lambda I_n) \mid \lambda \in \mathbb{F}_q^*\}$. Berger knew earlier in [5] that any linear isometry of the rank metric could be expressed as a multiplication by an element of $\mathbb{F}_{q^m}$ and an element of $\mathrm{GL}_n(F_q)$, but the structure was deduced by Morrison.

In addition to giving the group structure of $\mathrm{Lin}((\mathbb{F}_{q^m}^n, d_{\mathrm{R}}))$, Morrison classified the semi-linear and $\mathbb{F}_q$-linear rank metric isometries of $\mathbb{F}_{q^m}$.

Definition 3.1.14. Let V be a $\mathbb{K}$-vector space. A map $f\colon V \to V$ is called semi-linear if for every $\boldsymbol{v}, \boldsymbol{w} \in V$, and every $\lambda \in \mathbb{K}$, there exists $\nu \in \mathbb{K}$ such that

1. $f(\boldsymbol{v} + \boldsymbol{w}) = f(\boldsymbol{v}) + f(\boldsymbol{w})$,
2. $f(\lambda \boldsymbol{v}) = \lambda^\nu f(\boldsymbol{v})$.

Like linear isometries, semi-linear isometries also map subspaces to subspaces. More precisely we have the following.

Proposition 3.1.15. *Let $\mathcal{C}$ be a subspace of $\mathbb{F}_{q^m}^n$, and f a semi-linear isometry such that $f(\lambda \boldsymbol{v}) = \lambda^\nu f(\boldsymbol{v})$ for each $\boldsymbol{v} \in \mathbb{F}_{q^m}^n$. Then $f(\mathcal{C})$ is also a subspace of $\mathbb{F}_{q^m}^n$.*

Proof. Let $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k$ be a basis for $\mathcal{C}$. We have,

$$f\left(\sum_{i=1}^{k} a_i \boldsymbol{c}_i\right) = \sum_{i=1}^{k} a_i^\nu f(\boldsymbol{c}_i) \in \langle f(\boldsymbol{c}_1), \ldots, f(\boldsymbol{c}_k)\rangle_{\mathbb{F}_{q^m}}.$$

Thus,

$$f(\langle \boldsymbol{c}_1, \ldots, \boldsymbol{c}_k\rangle_{\mathbb{F}_{q^m}}) \subseteq \langle f(\boldsymbol{c}_1), \ldots f(\boldsymbol{c}_k)\rangle_{\mathbb{F}_{q^m}}, \tag{3.7}$$

and since f is an isometry, the cardinalities of the set of Equation (3.7) must be equal, and therefore $f(\mathcal{C})$ is a subspace of $\mathbb{F}_{q^m}^n$. □

We say that $\mathcal{C}_1$ is *semi-linearly equivalent* to $\mathcal{C}_2$ if there exists a semi-linear isometry, f, such that $f(\mathcal{C}_1) = \mathcal{C}_2$. From Lemma 3.1.8 we saw that the coordinate-wise Frobenius map is a semi-linear isometry of $\mathbb{F}_{q^m}^n$. Morrison showed that, together with the linear isometries, it generates the space of semi-linear isometries. Explicitly, the space of semi-linear isometries can be given by

$$\mathrm{Lin}((\mathbb{F}_{q^m}^n, d_{\mathrm{R}})) \rtimes \mathrm{Gal}(\mathbb{F}_{q^m}/\mathbb{F}_q), \tag{3.8}$$

and any semi-linear isometry can be realized by the composition of the linear isometries with the coordinate-wise Frobenius map.

If we consider only susbet of $\mathbb{F}_{q^m}^n$ which are $\mathbb{F}_q$-linear, then the semi-linear isometries above act on these spaces as $\mathbb{F}_q$-linear isometries. If we consider an element in $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ as an $m \times n$ matrix, $[\boldsymbol{x}]$, then the $\mathbb{F}_q$-linear maps can be represented in the following way [57].

Theorem 3.1.16. *Let $f\colon \mathbb{F}_{q^m}^n \to \mathbb{F}_{q^m}^n$ be an isometry of the rank metric such that f is an $\mathbb{F}_q$-linear map. Then, there exists $L \in \mathrm{GL}_m(\mathbb{F}_q)$, $M \in \mathrm{GL}_n(\mathbb{F}_q)$ such that*

$$f(\boldsymbol{a}) = \begin{cases} L[\boldsymbol{a}]M & \text{for all } \boldsymbol{a} \in \mathbb{F}_{q^m}^n \\ L[\boldsymbol{a}]^T M & \text{for all } \boldsymbol{a} \in \mathbb{F}_{q^m}^n, \end{cases}$$

where the latter can occur only if $m = n$.

If $\mathcal{C}_1$ and $\mathcal{C}_2$ are isometric by an $\mathbb{F}_q$-linear isometry, we will say that $\mathcal{C}_1$ and $\mathcal{C}_2$ are *sub-linearly equivalent*.

3.2 Code Constructions

While there are numerous examples of constructions of codes with efficient decoding algorithms for codes in the Hamming metric, the variety of code constructions is sparser in the case of rank metric codes. In this section, we will summarize the important known results for codes linear MRD codes–the generalized Gabidulin codes and more generally codes from Moore matrices [26], a twisted Gabidulin construction [70, 50], a new maximum rank distance code construction, and the low rank parity check codes [31].

There has been significant interest in the construction of maximum rank distance (MRD) codes which are not required to be $\mathbb{F}_{q^m}$-linear–Delsarte codes. Interesting achievements include a non-linear construction by Cossidente, Marino and Pavese [16], the establishment of a connection between Delsarte codes and semi-fields by Sheekey [70], and a generalization by Lunardon [50]. These constructions are in general not linear over $\mathbb{F}_{q^m}$, but in the case when they are in fact linear, there are no other known examples of $\mathbb{F}_{q^m}$-linear MRD codes. One of the important properties of $\mathbb{F}_{q^m}$-linear MRD codes is the fact that they can be presented far more efficiently than their counterparts which are only linear over a subfield. For example, a rank metric code of $\mathbb{F}_q$-dimension mk in $\mathbb{F}_{q^m}^n$ has cardinality q^{mk} and requires km basis elements to represent the space. Since each basis element belongs to $\mathbb{F}_{q^m}^n$, a presentation of the code would require $km^2n \log q$ bits. An $\mathbb{F}_{q^m}$-linear code of dimension k also has q^{mk} elements, however, requires only $kmn \log q$ bits to represent, thus saving a factor of m in the presentation size.

In Theorem 3.2.28 we present a construction of an MRD code which is linear over the alphabet $\mathbb{F}_{q^m}$ which is not semi-linearly equivalent to a generalized Gabidulin code.

3.2.1 MRD Codes

The classical way of defining Gabidulin codes is to view them as a q-analog of Reed-Solomon codes, that is, as the evaluation of q-polynomials. Denote the space of q-

linearized polynomials over $\mathbb{F}_{q^m}$ by

$$\mathbb{F}_{q^m}[x]^{(q)} = \left\{ \sum_i f_i x^{([i])} \mid f_i \in \mathbb{F}_{q^m} \right\}.$$

For $f \in \mathbb{F}_{q^m}[x]^{(q)}$, the q-degree of f is the power of the highest monomial term with non-zero coefficient. The subspace of polynomials of degree at most $[k]$ will be given by

$$\mathcal{L}_k^{(q)}(\mathbb{F}_{q^m}) = \{f \in \mathbb{F}_{q^m}[x]^{(q)} \mid \text{the } q\text{-degree of } f \text{ is at most } [k]\}.$$

Definition 3.2.1. Let $0 \leq k \leq n \leq m$, and $\boldsymbol{\gamma} = \{\gamma_0, \ldots, \gamma_{n-1}\} \subset \mathbb{F}_{q^m}$ be a set of elements which are independent over $\mathbb{F}_q$. The Gabidulin code, $\mathrm{Gab}_{n,k}(\boldsymbol{\gamma})$ is given by,

$$\mathrm{Gab}_{n,k}(\boldsymbol{\gamma}) = \{(f(\gamma_0), \ldots, f(\gamma_{n-1})) \mid f \in \mathcal{L}_k^{(q)}(\mathbb{F}_{q^m})\}.$$

The Singleton bound for codes in the Hamming metric also gives a bound for codes in the rank metric.

Theorem 3.2.2. *[26] Let $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ be a rank metric code with minimum rank distance d. Then,*

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

Proof. We can assume that $m \geq n$ or else consider the code in which all elements of $\mathcal{C}$ considered as matrices are transposed. Suppose that $\boldsymbol{x} \in \mathcal{C}$ is a codeword of minimum rank distance, $d \leq n$. Let $\sigma \in \mathrm{GL}_n(\mathbb{F}_q)$ be an isometry such that $\boldsymbol{x}\sigma = (x_1, \ldots, x_d, 0, \ldots, 0)$. $\mathcal{C}\sigma$ is then a code of minimum Hamming distance d, since any element with fewer than d non-zero coordinates would have minimum rank distance smaller than d, contradicting d the minimum rank distance of $\mathcal{C}$. Therefore, $\mathcal{C}\sigma$ must satisfy the Singleton bound for the Hamming distance, $\log_{q^m}(\mathcal{C}) \leq n - d + 1$. Since $\mathcal{C}\sigma$ is isometric to $\mathcal{C}$, we obtain the result. □

Definition 3.2.3. A code attaining the Singleton bound is called a *maximum rank distance (MRD)* code.

Since $x, x^{[1]}, ..., x^{[k-1]}$ form a basis for $\mathcal{L}_k^{(q)}(\mathbb{F}_{q^m})$, we can write down the canonical generator matrix for the Gabidulin code, $\mathrm{Gab}_{n,k}(\boldsymbol{\gamma})$, as

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_0^{[1]} & \gamma_1^{[1]} & \cdots & \gamma_{n-1}^{[1]} \\ & & \vdots & \\ \gamma_0^{[k-1]} & \gamma_1^{[k-1]} & \cdots & \gamma_{n-1}^{[k-1]} \end{pmatrix}. \tag{3.9}$$

Gabidulin codes were later generalized by Kshevetskiy and Gabidulin in [45] as follows.

Definition 3.2.4. Let $\boldsymbol{\gamma} = \{\gamma_0, \ldots, \gamma_{n-1}\} \in \mathbb{F}_{q^m}$ be linearly independent over $\mathbb{F}_q$ and $s \in \mathbb{N}$ be such that $\gcd(s, m) = 1$. The *generalized Gabidulin code* $\mathrm{GGab}_{n,k,s}(\boldsymbol{\gamma})$ is the linear code with generator matrix

$$\begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_0^{[s]} & \gamma_1^{[s]} & \cdots & \gamma_{n-1}^{[s]} \\ & & \vdots & \\ \gamma_0^{[s(k-1)]} & \gamma_1^{[s(k-1)]} & \cdots & \gamma_{n-1}^{[s(k-1)]} \end{pmatrix}.$$

It is well-known that the roots of $x^q - x$ in $\mathbb{F}_{q^m}$ are exactly the elements of $\mathbb{F}_q$. For our main results we need a generalization of this result and some preliminaries which can be found in [48].

Lemma 3.2.5. *If* $\gcd(s, m) = 1$*, then the roots in* $\mathbb{F}_{q^m}$ *of* $x^{[s]} - x$ *are exactly the elements of* $\mathbb{F}_q$*.*

Proof. Consider the field $F_{q^{ms}}$, so that both $\mathbb{F}_{q^m}$ and $\mathbb{F}_{q^s}$ can be viewed as subfields [48, Theorem 2.6]. Since m and s are coprime these two subfields only intersect in the base field $\mathbb{F}_q$. Moreover, the roots of $x^{[s]} - x$ in $\mathbb{F}_{q^{ms}}$ are exactly the elements of $\mathbb{F}_{q^s}$, hence the roots of it in $\mathbb{F}_{q^m}$ are the elements of $\mathbb{F}_q$. □

Corollary 3.2.6 ([45])**.** *Generalized Gabidulin codes are MRD codes.*

Proposition 3.2.7. *[18, 45]*

1. *Let* $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ *be an MRD code of dimension* k*. Then the dual code* $\mathcal{C}^\perp \subseteq \mathbb{F}_{q^m}^n$ *is an MRD code of dimension* $n - k$*.*

2. *The family of Generalized Gabidulin codes is closed under the action of taking duals. Specifically, if* $\gcd(s, m) = 1$*, then*

$$\mathrm{GGab}_{n,k,s}(\boldsymbol{\gamma})^\perp = \mathrm{GGab}_{n,n-k,s}(\boldsymbol{\beta}),$$

for some $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_n\}$ *a set of independent elements over* $\mathbb{F}_q$*.*

Generalized Gabidulin codes are closely related to Moore matrices which were introduced and studied in [56]. We can easily verify some of their elementary properties.

Definition 3.2.8. A matrix $M \in \mathbb{F}_{q^m}^{k \times n}$ is called a *Moore matrix* if there exists a $\boldsymbol{\alpha} \in \mathbb{F}_{q^m}^n$ such that row i of M is equal to $\boldsymbol{\alpha}^{([i-1])}$ for $i = 1, \ldots, k$. $\boldsymbol{\alpha}$ is called the generator of M.

Corollary 3.2.9. *Let* $M \in \mathbb{F}_{q^m}^{k \times n}$ *be a Moore matrix of column rank* $k \leq t \leq n$ *over* $\mathbb{F}_q$*. Then, we can write*

$$M = M'U, \tag{3.10}$$

with $M' \in \mathbb{F}_{q^m}^{k \times t}$ *a generator matrix of a Gabidulin code, and* $U \in \mathbb{F}_q^{t \times n}$*.*

Proof. Let columns $i_1, ..., i_t$ be independent. Then these columns form a Gabidulin code of length t. From Proposition 3.1.12 we obtain the result. □

Lemma 3.2.10. *Fix $1 \leq k \leq n \leq N$, and let $M \in \mathbb{F}_{q^m}^{k\times N}$ be a Moore matrix with generator $\boldsymbol{\alpha}$, where $\mathrm{rk}(\boldsymbol{\alpha}) = n \leq m$.*

1. *$\langle M\rangle_{\mathbb{F}_{q^m}}$ has dimension k, and minimum rank distance $n-k+1$.*
2. *If $0 < k < n$, then $\dim\left(\langle M\rangle_{\mathbb{F}_{q^m}} \cap \langle M\rangle_{\mathbb{F}_{q^m}}^{([1])}\right) = k-1$ and $\dim\left(\langle M\rangle_{\mathbb{F}_{q^m}} + \langle M\rangle_{\mathbb{F}_{q^m}}^{([1])}\right) = k+1$.*
3. *If $A \in \mathbb{F}_{q^m}^{k\times N}$ is another Moore matrix then $M+A$ is also a Moore matrix. Moreover, if the column rank of A is equal to $r < n-k+1$, then the minimum rank distance of $\langle M+A\rangle_{\mathbb{F}_{q^m}}$ is at least $n-k+1-r$.*
4. *If the minimum rank distance of $\langle M\rangle_{\mathbb{F}_{q^m}}$ is $d > 1$, then the minimum rank distance of $\langle M\rangle_{\mathbb{F}_{q^m}} + \langle M^{([1])}\rangle_{\mathbb{F}_{q^m}}$ is equal to $d-1$.*
5. *If the minimum rank distance of $\langle M\rangle_{\mathbb{F}_{q^m}}$ is $d > 1$, and $E \in \mathbb{F}_q^{N\times(N-s)}$ is a full rank matrix, then ME is a Moore matrix and the minimum rank distance of $\langle ME\rangle_{\mathbb{F}_{q^m}} = \langle M\rangle_{\mathbb{F}_{q^m}} E$ is at least $d-s$.*

Proof. 1. Let $\alpha_{i_1}, ..., \alpha_{i_n}$ be n independent coordinates of $\boldsymbol{\alpha}$. We observe that we can write $M = M'U$ as in Corollary 3.2.9, where $M' \in \mathbb{F}_{q^m}^{k\times n}$ is a generator of the Gabidulin code $\mathcal{G}_{n,k}((\alpha_{i_1}, ..., \alpha_{i_n}))$, in the form (3.9), and $U \in \mathbb{F}_q^{n\times N}$. From Corollary 3.2.9, we have

$$d_{\min}^{\mathrm{R}}(\langle M\rangle_{\mathbb{F}_{q^m}}) = \min\{d_{\min}^{\mathrm{R}}(\langle M'\rangle_{\mathbb{F}_{q^m}}), \mathrm{rk}(U)\} = \min\{n-k+1, n\} = n-k+1.$$

2. The first statement follows directly from the Moore matrix structure. It then follows that

$$\begin{aligned}\dim(\langle M\rangle_{\mathbb{F}_{q^m}} + \langle M^{([1])}\rangle_{\mathbb{F}_{q^m}}) &= \dim(\langle M\rangle_{\mathbb{F}_{q^m}}) + \dim(\langle M^{([1])}\rangle_{\mathbb{F}_{q^m}}) \\ &\quad - \dim(\langle M\rangle_{\mathbb{F}_{q^m}} \cap \langle M\rangle_{\mathbb{F}_{q^m}}^{([1])}) \\ &= k+1.\end{aligned}$$

3. The first statement follows from the fact that $(x+y)^{[i]} = x^{[i]} + y^{[i]}$ for any $x, y \in \mathbb{F}_{q^m}$. Therefore the Moore structure is preserved under addition of matrices. For the second part note that any element $\boldsymbol{a} \in \langle A\rangle_{\mathbb{F}_{q^m}}$ has rank at most r and any non-zero element $\boldsymbol{m}_i \in \langle M\rangle$ has rank at least $n-k+1$. The result follows from the reverse triangle inequality.

4. Since the minimum rank distance of $\langle M\rangle_{\mathbb{F}_{q^m}}$ is $d > 1$, it follows from (2) that $\dim\left(\langle M\rangle_{\mathbb{F}_{q^m}} + \langle M^{([1])}\rangle_{\mathbb{F}_{q^m}}\right) = k+1$. Then, from (1) we obtain that the minimum rank distance of $\langle M\rangle_{\mathbb{F}_{q^m}} + \langle M^{([1])}\rangle_{\mathbb{F}_{q^m}}$ is $n-(k+1)+1 = d-1$.

5. Let $E' \in \mathbb{F}_q^{N\times s}$ be such that $[E \mid E']$ has full rank. Then, $[E \mid E']$ is an isometry, and so $\langle M[E \mid E']\rangle_{\mathbb{F}_{q^m}}$ has minimum rank distance d. Removing the last s columns gives $\langle ME\rangle_{\mathbb{F}_{q^m}} = \langle M\rangle_{\mathbb{F}_{q^m}} E$, which can only decrease the rank by at most s.

□

We state the main result of Sheekey and Lunardon, which is the construction of the generalized twisted Gabidulin code.

Theorem 3.2.11. *[70, 50] Let $k \leq n \leq m$, $(s, m) = 1$, $h \in \mathbb{Z}_{\geq 0}$, and $\alpha_0, \ldots, \alpha_{n-1} \in \mathbb{F}_{q^m}$ be independent over $\mathbb{F}_q$. Suppose that $\eta \in \mathbb{F}_{q^m}$ satisfies $N_{q^{sm}/q^s}(\eta) \neq (-1)^{mk}$, and define the set*

$$\mathcal{H}_{k,s}(\eta, h) = \{a_0 x + a_1 x^{[s]} + \ldots + a_{k-1} x^{[s(k-1)]} + \eta a_0^{[h]} x^{[sk]} \mid a_i \in \mathbb{F}_{q^m}\}.$$

The code

$$\mathcal{C}_{k,s}(\eta, h) = \{(f(\alpha_0), f(\alpha_1), \ldots, f(\alpha_{n-1})) \mid f \in \mathcal{H}_{k,s}(\eta, h)\}$$

is an MRD code.

In general, $\mathcal{C}_{k,s}(\eta, h)$ is only $\mathbb{F}_q$-linear. However, when $\eta = 0$ we recover the standard definition of generalized Gabidulin codes and when $h = 0$, we obtain a family of codes which are linear over $\mathbb{F}_{q^m}$ and not generalized Gabidulin codes. In this case, one can obtain a generator matrix by evaluating the polynomials $x+\eta x^{[sk]}, x^{[s]}, \ldots, x^{[s(k-1)]}$ since they form a basis for $\mathcal{C}_{k,s}(\eta, 0)$ over $\mathbb{F}_{q^m}$. We can observe, then, that $\mathcal{C}_{k,s}(\eta, 0)$ always contains a generalized Gabidulin code of dimension $k - 1$ as a subcode.

3.2.2 New MRD Codes

In this section, we will focus only on $\mathbb{F}_{q^m}$-linear rank metric codes in $\mathbb{F}_{q^m}^n$ and so we drop the descriptor, $\mathbb{F}_{q^m}$-linear, and simply refer to them as MRD codes. We will present some preliminary results and observations, followed by constructions of new MRD codes, for small parameters. In this work we want to give new conditions for MRD codes and in particular retrieve an algebraic condition allowing one to efficiently determine if a given code is a generalized Gabidulin code. This will allow us to test if new constructions are generalized Gabidulin codes, or not. For this we will derive some properties for both MRD and generalized Gabidulin codes.

The following criterion for MRD codes was already given in [26]:

Proposition 3.2.12. *Let $H \in \mathbb{F}_{q^m}^{(n-k)\times n}$ be a parity check matrix of a rank metric code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$. Then $\mathcal{C}$ is an MRD code if and only if*

$$\mathrm{rk}(VH^T) = n - k,$$

for all $V \in \mathrm{Mat}_{(n-k)\times n}(\mathbb{F}_q)$ with $\mathrm{rk}(V) = n - k$.

This criterion is formulated with respect to the parity check matrix of a linear code. We can easily derive a criterion for the generator matrix of MRD codes from this:

Corollary 3.2.13. *Let $G \in \mathbb{F}_{q^m}^{k\times n}$ be a generator matrix of a rank-metric code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$. Then C is an MRD code if and only if*

$$\mathrm{rk}(VG^T) = k$$

for all $V \in \mathrm{Mat}_{k\times n}(\mathbb{F}_q)$, with $\mathrm{rk}(V) = k$.

Proof. The generator matrix, G, of $\mathcal{C}$ is a parity check matrix of the dual code $\mathcal{C}^\perp \subseteq \mathbb{F}_{q^m}^n$ of dimension $n-k$. It follows from Proposition 3.2.12 that $\mathcal{C}^\perp$ is an MRD code if and only if $\mathrm{rk}(VG^T) = k$ for all $V \in \mathbb{F}_q^{k\times n}$ with $\mathrm{rk}(V) = k$. From Proposition 3.2.7, $\mathcal{C}$ is MRD if and only if $\mathcal{C}^\perp$ is MRD. The statement follows. □

Lemma 3.2.14. *Any generator matrix $G \in \mathbb{F}_{q^m}^{k\times n}$ of an MRD code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ of dimension k has only non-zero maximal minors.*

Proof. Let $V = [\, I_k \mid 0 \,] \in \mathbb{F}_q^{k\times n}$. Then $\det(VG^T)$ is the maximal minor of G involving the first k columns. By Corollary 3.2.13, this minor is non-zero. Similarly, we can create all other maximal minors of G by multiplication by some other $V \in \mathrm{Mat}_{k\times n}(\mathbb{F}_q)$ of rank k. Thus, by Corollary 3.2.13 we obtain the statement of the lemma. □

Theorem 3.2.15. *Let $G \in \mathbb{F}_{q^m}^{k\times n}$ be a generator matrix of a rank-metric code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$. Then $\mathcal{C}$ is an MRD code if and only if for every $\sigma \in \mathrm{GL}_n(\mathbb{F}_q)$, each maximal minor of $G\sigma$ is non-zero.*

Proof. We first prove the converse direction. For this, let $\mathcal{C}$ be an MRD code. Then all elements in the orbit of $\mathcal{C}$ under $\mathrm{GL}_n(\mathbb{F}_q)$ are also MRD codes. Since $\mathrm{GL}_n(\mathbb{F}_q)$ acts on the columns of any generator matrix of $\mathcal{C}$, together with Lemma 3.2.14, we obtain that all maximal minors of any element of the form $G\sigma$ must be non-zero.

For the other direction, let $\mathcal{C}$ be a non-MRD code, so that there exists a non-zero codeword, $\boldsymbol{c} \in \mathcal{C}$, of rank at most $n-k$. Then, there exists $\sigma \in \mathrm{GL}_n(\mathbb{F}_q)$ such that

$$\boldsymbol{c}\sigma = \left(\begin{array}{ccc|ccc} 0 & \dots & 0 & * & \dots & * \end{array}\right),$$

where the first k coordinates are 0 and the last $n-k$ coordinates belong to $\mathbb{F}_{q^m}$. This in turn implies that there exists a generator matrix of $\mathcal{C}\sigma$ with $\boldsymbol{c}\sigma$ as a row. Then, the first maximal minor of any generator matrix for $\mathcal{C}\sigma$ will must be zero contradicting that $\mathcal{C}$ is an MRD code. □

As a remark, we note that similar result is known in the Hamming metric. A code, $\mathcal{C}$, in the Hamming metric is MDS if and only if every minor of a generator matrix for $\mathcal{C}$ is non-zero.

In order to be able to work with this criterion, we can slightly simplify it as follows. For this, denote by $\mathrm{UT}_n^*(\mathbb{F}_q)$ the subgroup of $\mathrm{GL}_n(\mathbb{F}_q)$ of upper triangular matrices with all ones on the diagonal.

Corollary 3.2.16. *Let $G \in \mathbb{F}_{q^m}^{k\times n}$ be a generator matrix of a rank-metric code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$. Then $\mathcal{C}$ is an MRD code if and only if for any $\sigma \in \mathrm{UT}_n^*(\mathbb{F}_q)$, every maximal minor of $G\sigma$ is non-zero.*

Proof. Note that $\mathrm{UT}_n^*(\mathbb{F}_q)$, together with the invertible diagonal matrices and the permutation matrices in $\mathrm{GL}_n(\mathbb{F}_q)$ generate the entire general linear group $\mathrm{GL}_n(\mathbb{F}_q)$. The action of the diagonal matrices multiplies the maximal minors of the generator matrix by a non-zero scalar, the action of the permutation matrices at most changes the sign of the maximal minors. Hence, these two subgroups do not change the non-zeroness of the maximal minors. Thus, we only need to consider the action of $\mathrm{UT}_n^*(\mathbb{F}_q)$. □

The following lemma generalizes Corollary 3.1.9.

Lemma 3.2.17. *Let $(s,m)=1$ and $\boldsymbol{v}=(v_1,\ldots,v_n)\in\mathbb{F}_{q^m}^n$ be of rank r over $\mathbb{F}_q$. Then $\boldsymbol{v},\boldsymbol{v}^{([s])},\ldots,\boldsymbol{v}^{([s(r-1)])}$ are linearly independent over $\mathbb{F}_{q^m}$.*

Proof. Assume that $\boldsymbol{v},\boldsymbol{v}^{([s])},\ldots,\boldsymbol{v}^{([s(r-1)])}$ are not linearly independent over $\mathbb{F}_{q^m}$, so that that there exist $\lambda_0,\ldots,\lambda_{r-1}\in\mathbb{F}_{q^m}$, at least one $\lambda_i\neq 0$, such that

$$\sum_{i=0}^{r-1}\lambda_i\boldsymbol{v}^{([is])}=0.$$

Then the q^s-linearized polynomial

$$p(x)=\sum_{i=0}^{r-1}\lambda_i x^{[si]}=\sum_{i=0}^{r-1}\lambda_i x^{(q^s)^i}\in\mathbb{F}_{q^{ms}}[x],$$

has $v_1,\ldots,v_n$ as roots. Since p is $\mathbb{F}_{q^s}$-linearized, all elements of the vector space $\langle v_1,\ldots,v_n\rangle_{\mathbb{F}_{q^s}}$ are also roots. Since $\langle v_1,\ldots,v_n\rangle_{\mathbb{F}_q}$ has dimension r, by [45, Lemma 4.3] also $\langle v_1,\ldots,v_n\rangle_{\mathbb{F}_{q^s}}$ has dimension r. Hence, there are q^{rs} roots of p in $\mathbb{F}_{q^{ms}}$. Hence p must have degree at least q^{rs}, which is a contradiction. □

The following straight-forward lemma is needed to prove Lemma 3.2.19.

Lemma 3.2.18. *Let $(s,m)=1$ and $\boldsymbol{w}_1,\ldots,\boldsymbol{w}_k\in\mathbb{F}_{q^m}^n$ be linearly independent over $\mathbb{F}_{q^m}$. Then $\boldsymbol{w}_1^{([s])},\ldots,\boldsymbol{w}_k^{([s])}\in\mathbb{F}_{q^m}^n$ are also linearly independent over $\mathbb{F}_{q^m}$*

Proof. Assume that $\boldsymbol{w}_1^{([s])},\ldots,\boldsymbol{w}_k^{([s])}$ are not linearly independent, so that there exist $\lambda_1,\ldots,\lambda_k\in\mathbb{F}_{q^m}$ with

$$\sum_{i=1}^{k}\lambda_i\boldsymbol{w}_i^{([s])}=0\Leftrightarrow\left(\sum_{i=1}^{k}\lambda_i^{[-s]}\boldsymbol{w}_i\right)^{([s])}=0\Leftrightarrow\sum_{i=1}^{k}\lambda_i^{[-s]}\boldsymbol{w}_i=0.$$

Thus, the vectors $\boldsymbol{w}_1,\ldots,\boldsymbol{w}_k$ are not linearly independent, a contradiction. □

The following result is a generalization of [35, Theorem 1].

Lemma 3.2.19. *Let $(s,m)=1$ and $W\subset\mathbb{F}_{q^m}^n$ be a subspace of dimension $k\leq n$ satisfying $W^{([s])}=W$. Then W has a generator matrix in $\mathbb{F}_q^{k\times n}$. In particular, W has a basis of elements of rank one over $\mathbb{F}_q$.*

Proof. If $\{\boldsymbol{w}_1, ..., \boldsymbol{w}_k\} \subset \mathbb{F}_{q^m}^n$, is a basis for W, then by Lemma 3.2.18 $\{\boldsymbol{w}_1^{([s])}, ..., \boldsymbol{w}_k^{([s])}\}$ is also a basis of W. Then, there exists an $A \in \mathrm{GL}_k(\mathbb{F}_{q^m})$ such that

$$\begin{pmatrix} w_{1,1}^{[s]} & w_{1,2}^{[s]} & \dots & w_{1,n}^{[s]} \\ w_{2,1}^{[s]} & w_{2,2}^{[s]} & \dots & w_{2,n}^{[s]} \\ & & \vdots & \\ w_{k,1}^{[s]} & w_{k,2}^{[s]} & \dots & w_{k,n}^{[s]} \end{pmatrix} = A \begin{pmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & w_{2,2} & \dots & w_{2,n} \\ & & \vdots & \\ w_{k,1} & w_{k,2} & \dots & w_{k,n} \end{pmatrix}.$$

Since the rightmost matrix has rank k, there exists a set of k linearly independent (over $\mathbb{F}_{q^m}$) columns. Without loss of generality, assume that the first k columns are linearly independent. Thus, the submatrix $W_1 = (w_{i,j})_{i,j=1}^{k}$ is invertible (and therefore $W_1^{([s])}$ is also invertible by Lemma 3.2.18), and so we can solve

$$A = W_1^{([s])} W_1^{-1}.$$

Define $W_2 = (w_{i,j})_{i=1\ j=k+1}^{k\ n}$. Then we have,

$$W_2^{([s])} = W_1^{([s])} W_1^{-1} W_2.$$

If we apply the coordinate-wise Frobenius map s times on both sides and use Lemma 3.1.8, we obtain,

$$\begin{aligned} W_2^{([2s])} &= W_1^{([2s])} (W_1^{-1})^{([s])} W_2^{([s])} \\ &= W_1^{([2s])} (W_1^{([s])})^{-1} W_1^{([s])} W_1^{-1} W_2 \\ &= W_1^{([2s])} W_1^{-1} W_2. \end{aligned}$$

Then, we have

$$W_1^{([2s])} (W_1^{-1})^{([s])} W_2^{([s])} = W_1^{([2s])} W_1^{-1} W_2.$$

Since $W_1^{([2s])}$ is invertible, we obtain

$$(W_1^{-1} W_2)^{([s])} = W_1^{-1} W_2,$$

and therefore we must have that $W_1^{-1} W_2$ has only entries in $\mathbb{F}_q$, by Lemma 3.2.5. Therefore, a generator matrix for W can be expressed as $W_1^{-1}[W_1 \mid W_2] = [I_k \mid W_1^{-1} W_2] \in \mathbb{F}_q^{k \times n}$. □

Proposition 3.2.20. *Let $(s, m) = 1$ and suppose that $\mathcal{C} \subset \mathbb{F}_{q^m}^n$ is a linear code of dimension $k \geq 2$ and minimum rank distance at least k. If $\dim(\mathcal{C} \cap \mathcal{C}^{[s]}) = k - 1$ (this automatically implies that $k < n$), then there exists a generator matrix for $\mathcal{C}$ of the form,*

$$G^* = \begin{pmatrix} g_1 & g_2 & \dots & g_n \\ g_1^{[s]} & g_2^{[s]} & \dots & g_n^{[s]} \\ & & \vdots & \\ g_1^{[s(k-1)]} & g_2^{[s(k-1)]} & \dots & g_n^{[s(k-1)]} \end{pmatrix},$$

where $g_1, \dots, g_n \in \mathbb{F}_{q^m}$.

Proof. We prove this inductively on k. First assume that $k = 2$. Then $\dim(\mathcal{C} \cap \mathcal{C}^{([s])}) = 1$, i.e. there exists $\boldsymbol{g}' \in \mathcal{C}$ such that $\mathcal{C} \cap \mathcal{C}^{([s])} = \langle \boldsymbol{g}' \rangle_{\mathbb{F}_{q^m}}$. Since $\boldsymbol{g}' \in \mathcal{C}^{([s])}$, we get that $\boldsymbol{g}'^{([-s])} \in \mathcal{C}$. The minimum rank distance of $\mathcal{C}$ is at least $k = 2$, i.e. the rank of $\boldsymbol{g}'^{([-s])}$ over $\mathbb{F}_q$ is at least 2. Then, by Lemma 3.2.17, $\boldsymbol{g}'^{([-s])}$ and $\boldsymbol{g}'$ are linearly independent. Hence they form a basis of $\mathcal{C}$ and we can rename $\boldsymbol{g} = \boldsymbol{g}'^{([-s])}$ to write a generator matrix as

$$G^* = \begin{pmatrix} \boldsymbol{g} \\ \boldsymbol{g}^{[s]} \end{pmatrix}.$$

We now explain the induction step $(k-1) \to k$. Let $W = \mathcal{C} \cap \mathcal{C}^{([s])}$. We know from Lemma 3.2.19 that $W^{([s])} \neq W$, because the minimum rank distance of $\mathcal{C}$ is at least $k \geq 2$. Since $W, W^{([s])} \subset \mathcal{C}^{([s])}$, both with codimension 1, we get $\langle W, W^{([s])} \rangle_{\mathbb{F}_{q^m}} = \mathcal{C}^{([s])}$. Then,

$$\dim(W \cap W^{([s])}) = \dim(W) + \dim(W^{([s])}) - \dim(W + W^{([s])}) = 2(k-1) - k = k - 2.$$

Furthermore, since $W \subset \mathcal{C}$, the minimum rank distance of W is at least k. Therefore, W satisfies the conditions of the induction hypothesis, and so we can express W in terms of some basis of the form

$$\{\boldsymbol{w}, \boldsymbol{w}^{([s])}, \ldots, \boldsymbol{w}^{([s(k-2)])}\}.$$

Hence, $\{\boldsymbol{w}, \boldsymbol{w}^{([s])}, \ldots, \boldsymbol{w}^{([s(k-2)])}\} \in \mathcal{C}$ and thus $\{\boldsymbol{w}^{([s])}, \boldsymbol{w}^{([2s])}, \ldots, \boldsymbol{w}^{([s(k-1)])}\} \in \mathcal{C}^{([s])}$. On the other hand, $\boldsymbol{w} \in W \subset \mathcal{C}^{([s])}$, i.e. $\{\boldsymbol{w}, \boldsymbol{w}^{([s])}, \ldots, \boldsymbol{w}^{([s(k-1)])}\} \in \mathcal{C}^{([s])}$. By Lemma 3.2.17 this set is linearly independent, i.e. it is a basis of $\mathcal{C}^{([s])}$. This in turn implies that $\{\boldsymbol{w}^{([-s])}, \boldsymbol{w}, \boldsymbol{w}^{([s])}, \ldots, \boldsymbol{w}^{([s(k-2)])}\}$ is a basis of $\mathcal{C}$. Define $\boldsymbol{g} = \boldsymbol{w}^{[-s]}$, and we obtain that $\{\boldsymbol{g}, \boldsymbol{g}^{([s])}, \ldots, \boldsymbol{g}^{([s(k-1)])}\}$ is a basis of $\mathcal{C}$. □

Lemma 3.2.21. *Let $\mathcal{C}$ be a linear MRD code of dimension $k < n$ with generator matrix*

$$G^* = \begin{pmatrix} g_1 & g_2 & \cdots & g_n \\ g_1^{[s]} & g_2^{[s]} & \cdots & g_n^{[s]} \\ & & \vdots & \\ g_1^{[s(k-1)]} & g_2^{[s(k-1)]} & \cdots & g_n^{[s(k-1)]} \end{pmatrix}.$$

Then $g_1, \ldots, g_n$ are linearly independent over $\mathbb{F}_q$.

Proof. We prove this by contradiction. Assume, without loss of generality, that g_1 is in $\langle g_2, ..., g_n \rangle_{\mathbb{F}_q}$, that is, that there exist $\lambda_2, \ldots, \lambda_n \in \mathbb{F}_q$ with $g_1 = \sum_{i=2}^n \lambda_i g_i$. Then

$$g_1^{[j]} = \left(\sum_{i=2}^n \lambda_i g_i \right)^{[j]} = \sum_{i=2}^n \lambda_i^{[j]} g_i^{[j]} = \sum_{i=2}^n \lambda_i g_i^{[j]}.$$

Therefore, $g_1^{[j]} \in \langle g_2^{[j]}, \ldots, g_n^{[j]} \rangle_{\mathbb{F}_q}$ for any $j \in \mathbb{N}$. Hence there exists $\sigma \in \mathrm{GL}_n(\mathbb{F}_q)$ such that the first column of $G^*\sigma$ is zero. It follows from Theorem 3.2.15 that $\mathcal{C}$ is then not an MRD code. □

Theorem 3.2.22. *If $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ is a linear MRD code of dimension k and $\dim(\mathcal{C} \cap \mathcal{C}^{[s]}) = k-1$, then $\mathcal{C}$ is a generalized Gabidulin code.*

Proof. If $k \leq (n+1)/2$, then the minimum distance of $\mathcal{C}$ is at least k. Then it follows from Proposition 3.2.20 that $\mathcal{C}$ has a generator matrix of the form

$$G^* = \begin{pmatrix} g_1 & g_2 & \cdots & g_n \\ g_1^{[s]} & g_2^{[s]} & \cdots & g_n^{[s]} \\ & & \vdots & \\ g_1^{[s(k-1)]} & g_2^{[s(k-1)]} & \cdots & g_n^{[s(k-1)]} \end{pmatrix}.$$

It follows from Lemma 3.2.21 that the g_i are linearly independent over $\mathbb{F}_q$. This is the definition of a generalized Gabidulin code.

If $k > (n+1)/2$, then it follows from Proposition 3.2.7 that the dual code $\mathcal{C}^\perp \subseteq \mathbb{F}_{q^m}^n$ has dimension $n-k$ and minimum distance $k+1 > n-k$, i.e. we can use Proposition 3.2.20 and Lemma 3.2.21 as before to show that $\mathcal{C}^\perp$ is a generalized Gabidulin code. Since the dual of a generalized Gabidulin code is again a generalized Gabidulin code, the statement follows. □

Corollary 3.2.23. *Let $\mathcal{C}$ be an MRD code which is not a generalized Gabidulin code. Then, $\mathcal{C}$ is not semi-linearly equivalent to a generalized Gabidulin code.*

Proof. First, we note that generalized Gabidulin codes are closed under semi-linear isometries. Therefore, if a code is not equal to a generalized Gabidulin, then it is also not semi-linearly equivalent to one. □

We are now in a position to construct new MRD codes which are not semi-linearly isometric to generalized Gabidulin codes.

Theorem 3.2.24. *All linear MRD codes in $\mathbb{F}_{q^m}^n$ of dimension $k = 1$ or $k = n-1$ are Gabidulin codes.*

Proof. Let $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ be an MRD code of dimension 1. Then the minimum rank distance is n and it can be generated by one vector in $\mathbb{F}_{q^m}^n$. Clearly this vector needs to have only entries that are linearly independent over $\mathbb{F}_q$, thus it is a Gabidulin code.

Since the dual of a Gabidulin code is again a Gabidulin code (see Proposition 3.2.7), the statement for codes of dimension $n-1$ follows. □

From here, we easily obtain the following:

Corollary 3.2.25. *All linear MRD codes of length $n \in \{1, 2, 3\}$ are Gabidulin codes.*

Lemma 3.2.26. *Any MRD code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ of dimension k has a generator matrix $G \in \mathbb{F}_{q^m}^{k \times n}$ in systematic form,*

$$G = [\, I_k \mid * \,].$$

Moreover, all entries of $*$ are from $\mathbb{F}_{q^m} \setminus \mathbb{F}_q$.

Proof. Assume that the generator matrix of $\mathcal{C}$ in reduced row echelon form has a row with pivot in column $i > k$. Then this row vector has at most $n-k$ many non-zero entries, which contradicts the minimum rank distance $n-k+1$ of $\mathcal{C}$. Therefore, all pivots are in columns $1, \ldots, k$, which proves the first statement. The second statement follows again from the minimum rank distance $n-k+1$ of the code, because every codeword needs to have at least $n-k$ entries from $\mathbb{F}_{q^m} \setminus \mathbb{F}_q$. □

In the first case not covered by Theorem 3.2.24, $n = m = 4$, $k = 2$ and $q = 2$, we can get the following statement, which was also proven, with quite different tools, in [70].

Proposition 3.2.27. *All linear MRD codes in $\mathbb{F}_{2^4}^4$ are Gabidulin codes.*

Proof. The case for codes of dimension $k = 1$ or $k = 3$ follows from Theorem 3.2.24. It remains to show the case $k = 2$. Then by Lemma 3.2.26 there exists a generator matrix of the form

$$G = \begin{pmatrix} 1 & 0 & a & b \\ 0 & 1 & c & d \end{pmatrix},$$

with $a, b, c, d \in \mathbb{F}_{2^4} \setminus \mathbb{F}_2$. For G to generate an MRD code, by Theorem 3.2.15 we have that

$$G \begin{pmatrix} 1 & u_1 & u_2 & u_3 \\ 0 & 1 & u_4 & u_5 \\ 0 & 0 & 1 & u_6 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & u_1 & u_2 + a & u_3 + au_6 + b \\ 0 & 1 & u_4 + c & u_5 + cu_6 + d \end{pmatrix}$$

needs to have only non-zero maximal minors for $u_1, \ldots, u_6 \in \mathbb{F}_2$. Thus we get the following inequations:

$$\begin{aligned} 1 &\neq 0 \\ u_4 + c &\neq 0 \\ u_5 + cu_6 + d &\neq 0 \\ (u_2 + a) - u_1(u_4 + c) &\neq 0 \\ (u_3 + au_6 + b) - u_1(u_5 + cu_6 + d) &\neq 0 \\ (u_2 + a)(u_5 + cu_6 + d) - (u_4 + c)(u_3 + au_6 + b) &\neq 0 \end{aligned}$$

Clearly the first inequation is always true; as is the second, since $u_4 \in \mathbb{F}_2$ and $c \notin \mathbb{F}_2$.

For G not to generate a Gabidulin code we need, from Theorem 3.2.22, that

$$\operatorname{rk} \begin{pmatrix} 1 & 0 & a & b \\ 0 & 1 & c & d \\ 1 & 0 & a^2 & b^2 \\ 0 & 1 & c^2 & d^2 \end{pmatrix} \neq 3.$$

Since $a, b, c, d \notin \mathbb{F}_2$ the rank of the above matrix is clearly at least 3. If the rank is equal to 4, then

$$(a^2 - a)(d^2 - d) - (b^2 - b)(c^2 - c) \neq 0.$$

Thus, including this with the equations we obtained earlier, we need to check that there is no solution to the system of inequations,

$$\begin{aligned} u_5 + cu_6 + d &\neq 0 \\ (u_2 + a) - u_1(u_4 + c) &\neq 0 \\ (u_3 + au_6 + b) - u_1(u_5 + cu_6 + d) &\neq 0 \\ (u_2 + a)(u_5 + cu_6 + d) - (u_4 + c)(u_3 + au_6 + b) &\neq 0 \\ (a^2 - a)(d^2 - d) - (b^2 - b)(c^2 - c) &\neq 0 \end{aligned}$$

for any $u_1, \ldots, u_6 \in \mathbb{F}_2$. With the help of a computer program one can check that there exist no solutions for $a, b, c, d \in \mathbb{F}_{2^4} \backslash \mathbb{F}_2$ for the above system of inequations. □

We can therefore try to construct non-Gabidulin MRD codes for $q > 2$ and $m \geq 4$.

Theorem 3.2.28. *Let $m > 4$, $\alpha \in \mathbb{F}_{q^m}$ a primitive element, and $\gamma \in \mathbb{F}_q$ a quadratic non-residue in $\mathbb{F}_q$ such that $\gamma \neq (\alpha^{[s]} + \alpha)^2$ for any $0 < s < m$ with $(s, m) = 1$. Then*

$$G = \begin{pmatrix} 1 & 0 & \alpha & \alpha^2 \\ 0 & 1 & \alpha^2 & \gamma\alpha \end{pmatrix}$$

is a generator matrix of an MRD code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^4$ of dimension $k = 2$ that is not a generalized Gabidulin code.

Proof. First we prove that $\mathcal{C}$ is an MRD code. For this, we use Corollary 3.2.16. Note that

$$\mathrm{UT}^*{}_4(\mathbb{F}_q) = \left\{ \begin{pmatrix} 1 & u_1 & u_2 & u_3 \\ 0 & 1 & u_4 & u_5 \\ 0 & 0 & 1 & u_6 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mid u_1, \ldots, u_6 \in \mathbb{F}_q \right\},$$

and

$$G \begin{pmatrix} 1 & u_1 & u_2 & u_3 \\ 0 & 1 & u_4 & u_5 \\ 0 & 0 & 1 & u_6 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & u_1 & u_2 + \alpha & u_3 + u_6\alpha + \alpha^2 \\ 0 & 1 & u_4 + \alpha^2 & u_5 + u_6\alpha^2 + \gamma\alpha \end{pmatrix}.$$

We need to show that all maximal minors of this matrix are non-zero for any values of $u_1, \ldots, u_6$. This gives the inequations:

$$\begin{aligned} 1 &\neq 0 \\ u_4 + \alpha^2 &\neq 0 \\ u_5 + \alpha^2 u_6 + \gamma\alpha &\neq 0 \\ (u_2 + \alpha) - u_1(u_4 + \alpha^2) &\neq 0 \\ (u_3 + \alpha u_6 + \alpha^2) - u_1(u_5 + \alpha^2 u_6 + \gamma\alpha) &\neq 0 \\ (u_2 + \alpha)(u_5 + \alpha^2 u_6 + \gamma\alpha) - (u_4 + \alpha^2)(u_3 + \alpha u_6 + \alpha^2) &\neq 0 \end{aligned}$$

One can easily see that the first four inequations are always true, since all u_i are in $\mathbb{F}_q$. We can rewrite the fifth inequation as

$$(u_1u_5 - u_3) + (u_1\gamma - u_6)\alpha + (u_1u_6 - 1)\alpha^2 \neq 0.$$

If the last term is zero then $u_1 = u_6^{-1}$. But then $u_1\gamma - u_6 = u_6^{-1}(\gamma - u_6^2) \neq 0$ because γ is a quadratic non-residue. Thus, in this case, the middle term of the above sum does not vanish, and so the inequation is always true. Lastly we can rewrite the sixth inequation as

$$(u_2u_5 - u_3u_4) + (u_2\gamma + u_5 - u_4u_6)\alpha + (u_2u_6 + \gamma - u_4 - u_3)\alpha^2 - \alpha^4 \neq 0.$$

This is always true, since the minimal polynomial of α has degree $m > 4$ and $u_1, \ldots, u_6, \gamma \in \mathbb{F}_q$, so nothing can cancel out the α^4-term.

It remains to prove that $\mathcal{C}$ is not a generalized Gabidulin code. For this we use Theorem 3.2.22 and compute

$$\begin{aligned}
\operatorname{rk}\begin{pmatrix} G \\ G^{[s]} \end{pmatrix} &= \operatorname{rk}\begin{pmatrix} 1 & 0 & \alpha & \alpha^2 \\ 0 & 1 & \alpha^2 & \gamma\alpha \\ 1 & 0 & \alpha^{[s]} & \alpha^{2[s]} \\ 0 & 1 & \alpha^{2[s]} & \gamma\alpha^{[s]} \end{pmatrix} \\
&= \operatorname{rk}\begin{pmatrix} 1 & 0 & \alpha & \alpha^2 \\ 0 & 1 & \alpha^2 & \gamma\alpha \\ 0 & 0 & \alpha^{[s]} - \alpha & \alpha^{2[s]} - \alpha^2 \\ 0 & 0 & \alpha^{2[s]} - \alpha^2 & \gamma(\alpha^{[s]} - \alpha) \end{pmatrix},
\end{aligned}$$

for any s with $\gcd(s, m) = 1$. Since $\alpha \notin \mathbb{F}_q$ this rank cannot be equal to 2, by Lemma 3.2.5. Hence, $\mathcal{C}$ is Gabidulin if and only if the determinant of the lower right submatrix from above is zero, i.e. if and only if

$$\begin{aligned}
&\gamma(\alpha^{[s]} - \alpha)^2 - (\alpha^{2[s]} - \alpha^2)^2 = 0 \\
&\Leftrightarrow \gamma(\alpha^{[s]} - \alpha)^2 = (\alpha^{2[s]} - \alpha^2)^2 \\
&\Leftrightarrow \gamma(\alpha^{[s]} - \alpha)^2 = (\alpha^{[s]} - \alpha)^2(\alpha^{[s]} + \alpha)^2 \\
&\Leftrightarrow \gamma = (\alpha^{[s]} + \alpha)^2.
\end{aligned}$$

This is not possible, due to the conditions on γ, which implies that $\mathcal{C}$ is not a generalized Gabidulin code. □

Example 3.2.29. Let $q = 3, m = 5$ and α a root of $x^5 + 2x^2 + x + 1$. Then $\gamma = 2$ is a quadratic non-residue in $\mathbb{F}_{3^5}$ and the code with generator matrix

$$G = \begin{pmatrix} 1 & 0 & \alpha & \alpha^2 \\ 0 & 1 & \alpha^2 & 2\alpha \end{pmatrix}$$

is an MRD code which is not a generalized Gabidulin code.

In Theorem 3.2.28, we were restricted to the case $m > 4$. We can find analog constructions for $m = 4$, as shown in the following examples. The proofs that these examples are also non-Gabidulin MRD codes is analogous to the one of Theorem 3.2.28, but when checking if the maximal minor of $G \cdot \mathrm{UT}_n^*(\mathbb{F}_q)$ involving the third and fourth column is non-zero we cannot use the argument that the minimal polynomial $m(x)$ of α has degree at least 4. Instead we need to write α^4 modulo $m(x)$ and show that the minor remains non-zero.

Example 3.2.30. Let $q = 3, m = 4$, and α a root of $x^4 - x^3 - 1$. Then $\gamma = 2$ is a quadratic non-residue in $\mathbb{F}_3$ and it fulfills the conditions that $\gamma \neq (\alpha^{[s]} + \alpha)^2$ for any $0 < s < m$ with $\gcd(s, m) = 1$. Now the code with generator matrix

$$G = \begin{pmatrix} 1 & 0 & \alpha & \alpha^2 \\ 0 & 1 & \alpha^2 & 2\alpha \end{pmatrix}$$

is an MRD but not a generalized Gabidulin code. To show that it is an MRD code we need to prove that the before mentioned minor is non-zero, i.e. that

$$(u_2u_5 - u_3u_4) + (2u_2 + u_5 - u_4u_6)\alpha + (u_2u_6 + 2 - u_4 - u_3)\alpha^2 - \alpha^4$$

$$\iff (u_2u_5 - u_3u_4 - 1) + (2u_2 + u_5 - u_4u_6)\alpha + (u_2u_6 + 2 - u_4 - u_3)\alpha^2 - \alpha^3$$

is non-zero for any $u_1, \ldots, u_6 \in \mathbb{F}_q$. This is clearly the case since nothing can cancel out the α^3-term.

Note that in the previous example we could have chosen any minimal polynomial of α that involves a non-zero term of order 3 (and a suitable γ). The same proof would then show that the generated code is MRD but not a generalized Gabidulin code.

We conclude with a final example over $\mathbb{F}_5$. A generalization for other values of q is straightforward.

Example 3.2.31. Let $q = 5, m = 4$, and α a root of $x^4 + x^3 + x^2 + x + 3$. Then $\gamma = 2$ is a quadratic non-residue in $\mathbb{F}_5$ and it fulfills the conditions that $\gamma \neq (\alpha^{[s]} + \alpha)^2$ for any $0 < s < m$ with $\gcd(s, m) = 1$. Now the code with generator matrix

$$G = \begin{pmatrix} 1 & 0 & \alpha & \alpha^2 \\ 0 & 1 & \alpha^2 & 2\alpha \end{pmatrix}$$

is an MRD but not a generalized Gabidulin code. To show that it is an MRD code we need to prove that the before mentioned minor is non-zero, i.e. that

$$(u_2u_5 - u_3u_4) + (u_2\gamma + u_5 - u_4u_6)\alpha + (u_2u_6 + \gamma - u_4 - u_3)\alpha^2 - \alpha^4$$

$$\Leftrightarrow (u_2u_5 - u_3u_4 + 2) + (u_2\gamma + u_5 - u_4u_6 + 4)\alpha + (u_2u_6 + \gamma - u_4 - u_3 + 4)\alpha^2 + 4\alpha^3$$

is non-zero for any $u_1, \ldots, u_6 \in \mathbb{F}_q$. This is clearly the case since nothing can cancel out the $4\alpha^3$-term.

From Corollary 3.2.23, these codes constructed are not semi-linearly equivalent to a generalized Gabidulin code. Recall from Theorem 3.1.16 that there are still other isometries of the rank metric. It is not so easy, on the other hand, to determine if the codes we constructed are sub-linearly equivalent to generalized twisted Gabidulin codes. Also, we note that is quite difficult to construct codes for larger values of n, k as the number of inequations grows rapidly. Nevertheless, these preliminary results indicate there may be other interesting classes of linear MRD codes that have not been explicitly constructed.

3.2.3 LRPC Codes

Low rank parity check (LRPC) codes were presented in [31] as a rank-metric analog to low density parity check (LDPC) codes in the Hamming metric.

Definition 3.2.32. Let $F \subset \mathbb{F}_{q^m}$ be an $\mathbb{F}_q$-subspace of dimension f. A code will be called a *low rank parity check code* with support F if there exists a parity check matrix such that all entries belong to F.

In an analogous way to LDPC codes, the structure of the parity check matrix makes these codes amenable to syndrome decoding. Let $H \in F^{(n-k)\times n}$ be a parity check matrix for an LDPC code with support $F = \langle F_1, \ldots, F_f \rangle_{\mathbb{F}_q}$, and suppose that $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}$ is a received vector with $\mathrm{rk}(\boldsymbol{e}) = r$. Let $E_1, \ldots, E_r$ be a basis for $\mathrm{supp}_{\mathrm{vs}}(\boldsymbol{e})$. Then each coordinate of

$$\boldsymbol{s} = H\boldsymbol{y}^T = H\boldsymbol{e}^T = \left(\sum_{i=1}^{n} H_{1,i}e_i, \ldots, \sum_{i=1}^{n} H_{n-k,i}e_i \right)^T$$

has entries which belong to the product space $\langle E.F \rangle = \langle E_i F_j \rangle_{\mathbb{F}_q}$. Let $S = \langle s_1, \ldots, s_{n-k} \rangle_{\mathbb{F}_q}$ for coordinate s_i of $\boldsymbol{s}$, and suppose that $S = \langle E.F \rangle$.

Consider

$$E' = \bigcap_{i=1}^{d} F_i^{-1} S. \tag{3.11}$$

For each $i = 1, \ldots, f$, $F_i^{-1}S$ necessarily contains E_j for $j = 1, \ldots, r$ and so $E \subset F_i^{-1}S$. Therefore, $E \subseteq E'$. It shown in [31] that with high probability we have

$$\mathrm{supp}_{\mathrm{vs}}(\boldsymbol{e}) = \bigcap_{i=1}^{d} F_i^{-1} S.$$

We can then write down equations for the error coordinates,

$$e_i = \sum_{j=1}^{r} e_{ij} E_j,$$

for each coordinate e_i of $\boldsymbol{e}$. Each coordinate of $\boldsymbol{s}$ belongs to the space $\langle E.F \rangle$ which has spanning set $\{E_i F_j\}_{i=1}^{r}{}_{j=1}^{f}$. We can write each coordinate s_ℓ of $\boldsymbol{s}$ as

$$s_\ell = \sum_{i=1}^{f} \sum_{j=1}^{d} e_{ij} E_i F_j,$$

which gives a system of rf variables and $n-k$ equations.

Theorem 3.2.33 ([31]). *Let $F \subset \mathbb{E}$ be a subspace of dimension f, and $H \in F^{(n-k)\times n}$ a parity check matrix for an LRPC code with support F. Then, we can decode a random error $\boldsymbol{e}$ of rank r such that $rf \leq n-k$ with failure probability $q^{-(n-k+1-rf)}$ and complexity $r^2(4f^2m + n^2)$.*

We will make several remarks about rank based cryptosystems based on LRPC codes at the end of Chapter 5.

Chapter 4

Coding-Based Cryptosystems

McEliece first proposed a cryptographic scheme based on ideas in coding theory in 1978. Despite having efficient encryption and decryption processes, at the time, it was considered infeasible for practical purposes and remained a theoretical curiosity. One of the problems is that the system requires a large public key size relative to number theoretic based cryptosystems with the same practical level of security. There has been a significant amount of work on reducing the public key size in recent years. The impetus behind this is the result of Shor, which would allow one to break RSA is polynomial time on a quantum computer [71]. Shortly after, Chuang et al. [13] generalized the algorithm to solve the DLP in elliptic curve groups, effectively breaking ECC with a quantum computer. However, it was shown in [20] that the proposal by McEliece resists Shor's algorithm and more generally algorithms based on coset sampling. These two results together have motivated a great deal of interest in post-quantum cryptography and in particular improvements in algorithms solving the general syndrome decoding problem [11, 8]. On modern computers and with new techniques, the original parameters proposed by McEliece have been found to be too small to be secure, but the idea to use Goppa codes has so far resisted structural attacks. Other codes, such as LDPC and Reed-Solomon codes have, on the other hand, been shown to be structurally insecure.

In Section 4.1 we give a brief introduction to the McEliece cryptosystem, which is the foundation for coding based cryptography. We also briefly outline the information set decoding attack. Section 4.2 outlines some of the most important rank-based cryptosystems. The first such, the GPT cryptosystem, was proposed by Gabidulin et al. in [28] and a generalization given later by Loidreau in [49] called the GGPT cryptosystem is given in Subsection 4.2.1. In Subsection 4.2.2, Overbeck's attack on the GPT cryptosystem is explained. The attack was used to cryptanalyze many of the existing variants in the literature, and several variants of the GGPT cryptosystem were proposed specifically to resist the attack. Another variant designed to resist Overbeck's attack called the column scrambler variant was proposed Gabidulin et al. [29] and is presented in Subsection 4.2.4.

4.1 McEliece Cryptosystem

Many constructions for linear codes exist and often their parameters can be found or at least bounded using a variety of techniques. However, the class of codes for which efficient decoding algorithms exist is much smaller; it is because the problem of decoding is known to be significantly difficult. Let $H \in \mathbb{F}_q^{(n-k)\times n}$ be a matrix and let $t \in \{0, \ldots, n\}$. The syndrome decoding problem, $\mathrm{SD}_q(H, t, \boldsymbol{s})$, is to determine if there exists an $\boldsymbol{x} \in \mathbb{F}_q^n$ such that $H\boldsymbol{x}^T = \boldsymbol{s}$ and $\mathrm{wt_H}(\boldsymbol{x}) = t$. Suppose that t is the weight of a correctable error, $\boldsymbol{e}$, of a code, $\mathcal{C}$, with parity check matrix H. Let H' be a parity check matrix for $\langle H\rangle_{\mathbb{F}_q}^{\perp} + \langle \boldsymbol{e}\rangle_{\mathbb{F}_q}$. Then the search version of $\mathrm{SD}_q(H', t, \mathbf{0})$ is equivalent to the problem of minimum distance decoding in $\langle H\rangle_{\mathbb{F}_q}^{\perp}$. Berlekamp et al. studied this problem in [7], and proved the following important result.

Theorem 4.1.1. *Let $H \in \mathbb{F}_{q^m}^{(n-k)\times n}$ be a random matrix of full rank and $t \in \{0, \ldots, n\}$. The problem $\mathrm{SD}_q(H, t, \boldsymbol{s})$ is* **NP***-complete.*

As a consequence of Theorem 4.1.1, efficient solutions exist mainly for special cases of the SD problem and tend to be for codes which are quite structured. Without knowledge of this structure, an attacker would have to resort to using generic attacks which would be relatively inefficient since the general problem is **NP**-complete. The most efficient generic attacks against the syndrome decoding problem are based on the idea of information set decoding [63].

We note that there are subproblems of the syndrome decoding problem which can be solved in polynomial time. For instance, if $t = 0$ the solution is trivial, and even if $t = 1$ a solution can be found in polynomial time by linear algebra.

The McEliece cryptosystem is fairly straight-forward.

Definition 4.1.2. Let $G \in \mathrm{GL}_k(\mathbb{F}_q)$ be a generator matrix of a code capable of correcting t errors, and `dec` a decoding algorithm with respect to G. Furthermore, choose matrices $S \in \mathrm{GL}_k(\mathbb{F}_q)$ and $\pi \in S_n \hookrightarrow \mathrm{GL}_n(\mathbb{F}_q)$. The public key for the system will be given by $(\kappa_{\mathrm{pub}}, t)$, where

$$\kappa_{\mathrm{pub}} = SG\pi,$$

and the private key by

$$\kappa_{\mathrm{pvt}} = (S, \pi, \mathtt{dec}).$$

Such a system is called a *McEliece* cryptosystem.

Suppose that Bob wants to send a message $\boldsymbol{m} \in \mathbb{F}_q^k$ to Alice. Using Alice's public key, Bob chooses a random error $\boldsymbol{e}$ of weight at most t, and sends

$$\boldsymbol{y} = \boldsymbol{m}SG\pi + \boldsymbol{e}.$$

Alice then computes $\boldsymbol{y}\pi^{-1} = \boldsymbol{m}SG + \boldsymbol{e}\pi^{-1}$. Since the weight of $\boldsymbol{e}\pi^{-1}$ is unchanged by a permutation matrix, this error also has weight at most t. Therefore, Alice can decode to

$$\mathtt{dec}(\boldsymbol{y}\pi^{-1}) = \boldsymbol{m}S.$$

Inverting S, Alice can then obtain $\boldsymbol{m}$.

The security is based on the assumption that $SG\pi$ will appear as a random matrix to an eavesdropper. Without knowledge of the components of the private key, it will then be difficult to break the system. An attack on a PKC cryptosystem has one of two forms. The first are *structural* attacks–those which attempt to recover some information about the private key from the public key–and *generic* attacks–those which attempt to solve the SD problem. Sidel'nikov and Shestakov, for instance, managed to break the McEliece cryptosystem when the underlying code is Reed-Solomon by exploiting their predictable structure [72]. On the other hand, information set decoding is a generic attack–it can be applied to an arbitrary linear code.

Generic attacks provide an upper bound for the security of a cryptosystem. As a consequence, they also bound the parameters required for these systems to be considered secure. Unfortunately, because of the relative efficiency of generic attacks against the syndrome decoding problem, the parameters of McEliece cryptosystems must be quite large.

Prange in [63] introduced the idea of information set decoding upon which the fastest algorithms for solving the SD problem are based. The idea is relatively straight forward. Let $G \in \mathbb{F}_q^{k\times n}$ be a generator matrix for a code capable of correcting t errors. Given an error, $\boldsymbol{e} \in \mathbb{F}_q^n$, if one can guess k coordinates of $\boldsymbol{e}$ which are 0, then the message can be recovered from the submatrix of G whose columns correspond to the zero coordinates of $\boldsymbol{e}$. There have been several refinements of this idea an improvements in algorithms with the best known algorithm for solving the SD problem due to Becker et al. achieving a complexity of approximately $2^{n/20}$ [25, 21, 4].

4.2 Cryptosystems Based on Rank Metric Codes

4.2.1 GPT and GGPT Cryptosystem

Definition 4.2.1. Let $S \in \mathrm{GL}_k(\mathbb{F}_{q^m})$, $G \in \mathbb{F}_{q^m}^{k\times n}$ be a generator matrix of a Gabidulin code, say $\mathrm{Gab}_{n,k}(\boldsymbol{\alpha})$, capable of correcting t' errors, and $X \in \mathbb{F}_{q^m}^{k\times n}$ be a matrix of column rank $t < t'$. Let

$$G_{\mathrm{pub}} = SG + X. \tag{4.1}$$

A *GPT cryptosystem* is one in which the public key is given by the pair

$$\kappa_{\mathrm{pub}} = (G_{\mathrm{pub}}, t' - t), \tag{4.2}$$

and the private key is given by

$$\kappa_{\mathrm{pvt}} = (G, S). \tag{4.3}$$

An encryption of a message $\boldsymbol{m} \in \mathbb{F}_{q^m}^k$ is given by

$$\boldsymbol{m}G_{\mathrm{pub}} + \boldsymbol{e} = \boldsymbol{m}SG + \boldsymbol{m}X + \boldsymbol{e},$$

where $\boldsymbol{e} \in \mathbb{F}_{q^m}^n$ is a randomly chosen vector of rank at most $t' - t$. The product $\boldsymbol{m}S$ can be recovered from a decoding algorithm for $\mathrm{Gab}_{n,k}(\boldsymbol{\alpha})$ because all elements of $\langle X\rangle_{\mathbb{F}_{q^m}}$

have weight at most t. Specifically, if $\mathrm{wt}_{\mathrm{R}}(\boldsymbol{e}) \leq t' - t$,

$$\mathrm{wt}_{\mathrm{R}}(\boldsymbol{m}X + \boldsymbol{e}) \leq \mathrm{wt}_{\mathrm{R}}(\boldsymbol{m}X) + \mathrm{wt}_{\mathrm{R}}(\boldsymbol{e}) \leq t'.$$

Inverting S, the message $\boldsymbol{m}$ can then be recovered. We will call the elements of the form $\boldsymbol{m}X$ the *designed error* associated with the encryption of $\boldsymbol{m}$, and X the *designed error matrix*.

In [64, 49] the authors consider an alternative version which we call the *generalized GPT (GGPT) cryptosystem*. This system uses a public matrix of the form

$$\hat{G}_{\mathrm{pub}} = S[X \mid G]\sigma \in \mathbb{F}_{q^m}^{k \times (n+\hat{t})}, \tag{4.4}$$

where $G \in \mathbb{F}_{q^m}^{k \times n}$ is a generator matrix for a Gabidulin code, $X \in \mathbb{F}_{q^m}^{k \times \hat{t}}$, $S \in \mathrm{GL}_k(\mathbb{F}_{q^m})$, and $\sigma \in \mathrm{GL}_{n+\hat{t}}(\mathbb{F}_q)$. The public key is given by

$$\kappa_{\mathrm{pub}} = (\hat{G}_{\mathrm{pub}}, t'), \tag{4.5}$$

and the private key is given by

$$\kappa_{\mathrm{pvt}} = (G, S, \sigma). \tag{4.6}$$

In the GGPT cryptosystem, an encryption of $\boldsymbol{m} \in \mathbb{F}_{q^m}^k$ is given by

$$\boldsymbol{m}\hat{G} + \boldsymbol{e},$$

with $\mathrm{rk}(\boldsymbol{e}) \leq t'$. To recover $\boldsymbol{m}$, one first computes

$$(\boldsymbol{m}\hat{G} + \boldsymbol{e})\sigma^{-1},$$

and then ignores the first $\hat{t}$ coordinates. The last n coordinates of $\boldsymbol{e}\sigma^{-1}$ have weight at most t', and therefore decoding the last n coordinates with respect to G, one obtains $\boldsymbol{m}S$, and by applying S^{-1}, the message $\boldsymbol{m}$ can be recovered.

4.2.2 Overbeck's Attack

We will describe Overbeck's attack from [62] for the case of the GGPT cryptosystem; the attack for the GPT case is analogous. Let $G \in \mathbb{F}_{q^m}^{k \times n}$ be a generator matrix for the Gabidulin code, $\mathrm{Gab}_{n,k}(\boldsymbol{\alpha})$ and X a matrix of rank $\hat{t}$. The first step in Overbeck's attack is to consider the extended matrix

$$G_{\mathrm{ext}} = \begin{pmatrix} S[X \mid G]\sigma \\ (S[X \mid G]\sigma)^{([1])} \\ \vdots \\ (S[X \mid G]\sigma)^{([n-k-1])} \end{pmatrix} = \tilde{S} \left(\begin{array}{c|c} X & G \\ X^{([1])} & G^{([1])} \\ \vdots & \vdots \\ X^{([n-k-1])} & G^{([n-k-1])} \end{array}\right) \sigma,$$

where $\tilde{S} = \mathrm{diag}(S, S^{([1])}, \ldots, S^{([n-k-1])})$ is a block diagonal matrix. Since the n rightmost columns of $G_{\mathrm{ext}}\sigma^{-1}$ span the Gabidulin code $\mathrm{Gab}_{n,n-1}(\boldsymbol{\alpha})$, the matrix can be brought into the form of

$$G'_{\mathrm{ext}} = \tilde{S}' \left(\begin{array}{c|c} X^* & G^* \\ X^{**} & 0 \end{array}\right) \sigma, \tag{4.7}$$

by some suitable row transformation, where $X^* \in \mathbb{F}_{q^m}^{(n-1)\times \hat{t}}$, $G^* \in \mathbb{F}_{q^m}^{(n-1)\times n}$ is a generator matrix of $\mathrm{Gab}_{n,n-1}(\boldsymbol{\alpha})$, and $X^{**} \in \mathbb{F}_{q^m}^{(k-1)(n-k-1)\times \hat{t}}$. If X^{**} has rank $\hat{t}$, then any element of $\langle G'_{\mathrm{ext}} \rangle^{\perp} = \langle G_{\mathrm{ext}} \rangle^{\perp}$ has the form $[0 \mid \boldsymbol{h}](\sigma^{-1})^T$, where $\boldsymbol{h} \in \mathrm{Gab}_{n,n-1}(\boldsymbol{\alpha})^{\perp}$.

Because $(\sigma^{-1})^T$ is an isometry, the vector $[0 \mid \boldsymbol{h}](\sigma^{-1})^T$ has rank n. Therefore, one can find a $\rho \in \mathrm{GL}_{n+\hat{t}}(\mathbb{F}_q)$ and $\boldsymbol{h}^* \in \mathbb{F}_{q^m}^n$ such that

$$[0 \mid \boldsymbol{h}](\sigma^{-1})^T = [0 \mid \boldsymbol{h}^*]\rho.$$

Any such ρ must be of the form

$$\rho = \left(\begin{array}{c|c} A & B \\ \hline 0 & C \end{array}\right)(\sigma^{-1})^T,$$

where $A \in \mathrm{GL}_{\hat{t}}(\mathbb{F}_q), B \in \mathbb{F}_q^{\hat{t}\times n}$, and $C \in \mathrm{GL}_n(\mathbb{F}_q)$. We can then observe that

$$S[X \mid G]\sigma\rho^T = [SXA^T + SGB^T \mid SGC^T].$$

The matrix SGC^T is a generator matrix for a Gabidulin code. For an encrypted message, $\boldsymbol{y}$, we compute $\boldsymbol{y}\rho^T$ and ignore the first $\hat{t}$ coordinates. Decoding the remaining n coordinates with respect to SGC^T as in Lemma 5.2.1 we can decrypt the message.

4.2.3 GGPT Variants

Overbeck's attack succeeds because we can recover a one-dimensional space as the kernel of G'_{ext}, from (4.7). This space contains the information to reconstruct a suitable unscrambling matrix, ρ. If the rank of X^{**} is smaller than $\hat{t}$, then we cannot be guaranteed that $\langle X^* \rangle_{\mathbb{F}_{q^m}} \subset \langle X^{**} \rangle_{\mathbb{F}_{q^m}}$. The attack can be modified by enumerating over all possible one-dimensional subspaces of the kernel until an element of the form $[0 \mid \boldsymbol{h}](\sigma^{-1})^T$, for $\boldsymbol{h} \in \mathrm{Gab}_{n,n-1}(\boldsymbol{\alpha})$ is found. The probability of this occurance is that of finding a particular one-dimensional subspace in a $\dim\left(\langle G_{\mathrm{ext}} \rangle^{\perp}_{\mathbb{F}_{q^m}}\right)$-dimensional space. If X^{**} is sufficiently rank deficient, then this problem becomes computationally infeasible. However, when X is randomly chosen, it is highly probable that X^{**} has full rank.

Two strategies emerged in order to exploit this. The first is the strategy considered in [64] which is to restrict X by some design specifications so that X^{**} can be forced to be rank deficient. The second, presented in [49], is to use a randomly chosen designed error matrix X of low rank, a, and high column rank, $\hat{t}$, so that the rank of X^{**} can be bounded from above. Note that $\mathrm{colrk}_{\mathbb{F}_q}(X) \leq \min\{\hat{t}, am\}$, and a random matrix of rank a will achieve this value with high probability. Here, it is assumed that the smaller value is equal to $\hat{t}$ (so that $\mathrm{colrk}_{\mathbb{F}_q}(X) = \hat{t}$), so that if we choose $\hat{t} > a(n-k)$, then

$$X_{\mathrm{ext}} = \begin{pmatrix} X \\ X^{([1])} \\ \vdots \\ X^{([n-k-1])} \end{pmatrix}$$

has rank at most $a(n-k)$, and column rank $\hat{t}$. Therefore,

$$\mathrm{rk}(X^{**}) \leq \mathrm{rk}(X_{\mathrm{ext}}) \leq a(n-k) < \hat{t}.$$

We make the following a definition.

Definition 4.2.2. A GGPT system of the form (4.4) where the designed error matrix is a randomly chosen matrix of rank a and column rank $\hat{t}$ satisfying $a(n-k) < \hat{t}$ will be called the *LGGPT* variant.

We note that the restriction on $\hat{t}$ necessitates an increase in the public key size. In particular, the key size of the LGGPT variant is on the order of $(m \log q)k(n+\hat{t}) > (m \log q)k(n + a(n-k))$ bits.

The approach taken in [64] is to algebraically design X so that X^{**} has a predictable rank deficit. We will denote the ith row of X by X_i, and $\boldsymbol{g}$ the generating element of the underlying Gabidulin code. The extended public matrix, can be put in the form

$$G_{\mathrm{pub,ext}} = \tilde{S} \left(\begin{array}{c|c} X_1 & \boldsymbol{g} \\ \vdots & \vdots \\ X_k & \boldsymbol{g}^{([k-1])} \\ X_k^{([1])} & \boldsymbol{g}^{([k])} \\ \vdots & \vdots \\ X_k^{([n-k-1])} & \boldsymbol{g}^{([n-2])} \\ X' & 0 \\ \vdots & \vdots \\ (X')^{([n-k-2])} & 0 \end{array} \right),$$

where

$$X' = \begin{pmatrix} X_1^{([1])} - X_2 \\ \vdots \\ X_{k-1}^{([1])} - X_k \end{pmatrix}.$$

Now, if we let $X = X_{\mathrm{Moore}} + Z$ where X_{Moore} is a Moore matrix and Z has column rank a, then we note the Moore component of X will cancel in X', leaving

$$X' = \begin{pmatrix} Z_1^{([1])} - Z_2 \\ \vdots \\ Z_{k-1}^{([1])} - Z_k \end{pmatrix}.$$

Since Z has column rank a, then X' has column rank at most $a < \hat{t}$. Since the coordinate-wise Frobenius map preserves the Grassmann support of X' by Lemma 3.1.8, we have

$$\dim \left(\sum_{i=0}^{n-k-2} \langle X' \rangle_{\mathbb{F}_{q^m}} \right) = a,$$

and therefore Overbeck's attack can not be applied. In the paper where this idea was proposed, the system was referred to as the *smart approach*. Therefore, we make the following definition.

Definition 4.2.3. The GGPT system in which the designed error matrix is of the form

$$X = X_{\text{Moore}} + Z,$$

where $\text{colrk}_{\mathbb{F}_q}(Z) = a < \hat{t}$ will be called the *SA* variant.

Comparing the LGGPT and SA variants, we note that the SA variant should be able to be designed for significantly smaller parameter choices than the LGGPT system. We will cryptanalyze both, in Chapter 5.

4.2.4 Column Scrambler Variant

In the original McEliece cryptosystem, the generator matrix is hidden by a row scrambling matrix and a column permutation matrix. Importantly for the decryption step, the column permutation matrix is an isometry in the Hamming metric, and hence fixes the weight of the error vector which is added by the sender. For Gabidulin codes, the semi-linear isometries of the rank metric do not disguise the structure of the code enough. For instance, one could recover a decoding algorithm by Lemma 5.2.1.

It was proposed in [29] to use column scrambling matrices which are not isometries, in order to disguise a secret generator matrix. Of course, not any element of $\text{GL}_n(\mathbb{F}_{q^m})$ can be used; it is proposed to take a family, $\mathcal{P}$, of matrices with the property that for any $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ and any element $P \in \mathcal{P}$, $|\text{rk}(\boldsymbol{x}) - \text{rk}(\boldsymbol{x}P)|$ is bounded. Then, the error vector chosen at the sender increases in rank in a predictable way during the decryption process. A similar idea is considered in [3] for codes in the Hamming metric. We note that this is not to be confused with the column scrambler variant of Ourivski et al. in [60], which was broken by Overbeck.

Let G be a Gabidulin code and t be the error correction capability of $\langle G \rangle_{F_{q^m}}$. Let $0 < t_1 < t$ be a design parameter and let $\mathcal{P}_{n,t,t_1}(\mathbb{F}_{q^m}) \subset \text{GL}_n(\mathbb{F}_{q^m})$ be given by

$$\begin{aligned}\mathcal{P}_{n,t,t_1} = \{[Q_1 \mid Q_2]\sigma \mid & Q_1 \in \mathbb{F}_{q^m}^{n\times(t-t_1)}, Q_2 \in \mathbb{F}_q^{n\times(n-t+t_1)} \\ & \text{s.t. } \text{rk}([Q_1 \mid Q_2]) = n, \sigma \in \text{GL}_n(\mathbb{F}_q)\}.\end{aligned}$$

For any $\boldsymbol{e} \in \mathbb{F}_{q^m}^n$ of rank at most t_1, and $P = [Q_1 \mid Q_2]\sigma \in \mathcal{P}_{n,t,t_1}$ we have

$$\begin{aligned}\text{rk}(\boldsymbol{e}P) &= \text{rk}([\boldsymbol{e}Q_1 \mid \boldsymbol{e}Q_2]\sigma) \\ &= \text{rk}([\boldsymbol{e}Q_1 \mid \boldsymbol{e}Q_2]) \\ &\leq \text{rk}([\boldsymbol{e}Q_1 \mid 0]) + \text{rk}([0 \mid \boldsymbol{e}Q_2]) \\ &\leq t - t_1 + t_1 = t.\end{aligned}$$

Definition 4.2.4. Let $G \in \mathbb{F}_{q^m}^{k\times n}$ be a t-error correcting Gabidulin code. The column scrambler (CS) variant is one in which the public key is of the form

$$\kappa_{\text{pub}} = (SGP^{-1}, t_1), \tag{4.8}$$

for $t_1 < t$, and $P \in \mathcal{P}_{n,t,t_1}$. The private key is given by S^{-1}, G, and P.

The encryption of a message $\boldsymbol{m} \in \mathbb{F}_{q^m}^k$ is performed by simply encoding $\boldsymbol{m}$ with the public key and adding a random rank error,

$$\boldsymbol{m}SGP^{-1} + \boldsymbol{e},$$

for a randomly chosen $\boldsymbol{e}$ of rank weight at most t_1. Decryption proceeds by applying P, yielding $\kappa_{\text{pub}}P = \boldsymbol{m}SG + \boldsymbol{e}P$. Then the receiver decodes with respect to G, obtaining $\boldsymbol{m}S$, since $\text{rk}(\boldsymbol{e}P) \leq t$. Applying S^{-1}, the receiver can correctly recover $\boldsymbol{m}$.

Overbeck's attack fails against the CS variant because the isometry used to disguise the code is no longer invariant under the coordinate-wise Frobenius map [64]. The security of this system is therefore based on the idea that the inverse of the matrices in $\mathcal{P}$ will appear random and therefore disguise the matrix G. We will cryptanalyze this system in Section 4.2.4.

Chapter 5

Attacks on Rank-based Cryptosystems

The suggestion that codes in the rank metric could be safer to use than codes in the Hamming metric was motivated by preliminary results concerning combinatorial solutions for solving the rank syndrome decoding (RSD) problem–the analogous problem to the SD problem in the rank metric [61, 12]. These results indicated that to solve the RSD problem was significantly more difficult than the SD problem. As a consequence, one could potentially make cryptosystems based on codes in the rank metric with smaller parameters than that which would be feasible in the Hamming metric. The problem considered is actually a subproblem of the MinRank problem, which is known to be **NP**-complete [10]. However, as a sub-problem, the RSD problem is not known to be **NP**-complete.

Part of the difficulty in finding efficient algorithms for the RSD problem is that the structure of errors in the rank metric is not amenable to the same techniques–notably information set decoding–in which the fastest algorithms for solving the SD problem are based. As an illustration of why it is believed that the RSD problem is also very difficult, recall that $\mathcal{B}_t^{\mathrm{H}}(\mathbf{0}) \subset \mathcal{B}_t^{\mathrm{R}}(\mathbf{0})$. Therefore, if we can solve the rank syndrome decoding problem for $\mathcal{C}$ with respect to the rank distance, this also gives an algorithm (for the same t) with respect to the Hamming distance. While this would only solve a subproblem of the SD problem, a polynomial time solution would be surprising.

Rank metric codes were first suggested for use in cryptography by Gabidulin, Paramanov, and Trejakov (GPT) [28], and since then a number of proposals for systems have arisen [60, 27, 6]. One key aspect of many of these systems is their use of a Gabidulin code underpinning the cryptosystem. As a consequence, Gibson proposed the first structural attack which Overbeck extended, managing to cryptanalyze most of these systems simultaneously [34, 62]. In light of these structural flaws the GPT cryptosystem in its original form–as well as many variants–is infeasible, however, in the wake of the attack, several modifications were made in order to explicity resist Overbeck's attack [49, 64, 29]. One approach was to consider a generalized GPT (GGPT) system in which one embeds the structure of the Gabidulin code into a larger space, and therefore can have more complicated designed error matrices. In this direction, two cryptosystems stand out–one proposed by Loudreau in [49] and another proposed by Rashwan et al. in [64]. A sepa-

rate approach in [29] is to use non-isometric transformations of Gabidulin-based codes, similar to the approach adopted in [3]. In this chapter we introduce a new type of attack based on analyzing the elements of rank one in an extended public key. Using this new strategy, we can cryptanalyze all three systems, effectively breaking them. Additionally, our attack generalized Overbeck's attack for the original GPT system and manages to extend the attack to break the GPT system for all parameters.

The main results of this chapter can be found in [40]. In Section 5.1, we give a short introduction to the MinRank and RSD problems as well as well as a summary of the most efficient generic attacks. In Section 5.2, we prove some basic results that we will need for our cryptanalysis. In Section 5.3 we give a new attack which can be used to break all possible parameter sets for the GPT cryptosystem; generalizing Overbeck's attack. In Section 5.4, we show how our attack allows us to extend Overbeck's attack to variants of the designed to resist Overbeck's attack. In two cases we are able to entirely break the system, and in the other we are able to establish parameter values for which the system is not safe. We note that some of the parameters for these systems are already insecure based on generic attacks, however, our attacks are polynomial and therefore more resilient to changes in parameter values [30]. Lastly, we use some observations from the previous attacks to generalize the polynomial-time attack criterion by Gaborit at the end of Section 5.5. The results of Section 5.5 are being finalized and will be presented shortly.

5.1 MinRank and RSD Problems

Let R be a commutative ring with subsets $S, E \subseteq R$ and let $M \in E[x_1, \ldots, x_t]^{m \times n}$. Fix an integer, $r \geq 0$. The MinRank problem is to determine if there exists a solution so that

$$\min_{(a_1,\ldots,a_t)\in S^t} \mathrm{rk}\left(M(a_1, \ldots, a_t)\right) \leq r.$$

The MinRank problem was proposed by Buss et al. in [10], in which it was shown that the complexity of the problem depends heavily upon R and S. For our purposes, we are interested in the case when $E = S = \mathbb{F}_q$ is a finite field. In this case, it was shown that the MinRank problem is **NP**-complete.

Suppose that $\mathcal{C} \subset \mathbb{F}_{q^m}^n$ is an $\mathbb{F}_q$-linear rank metric code with basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_t$. Set $M_i = [\boldsymbol{b}_i]$ where $[\boldsymbol{b}_i] \in \mathbb{F}_q^{m\times n}$ is an expansion of $\boldsymbol{b}_i$ into a matrix whose columns are the elements of $\boldsymbol{b}_i$ expanded according to a basis of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ in the canonical way. If $\mathcal{C}$ is used to encrypt a message in the form

$$\boldsymbol{y} = \sum_{i=1}^{t} x_i \boldsymbol{b}_i + \boldsymbol{e},$$

where $\mathrm{rk}(\boldsymbol{e}) \leq \lfloor (d_{\min}^{\mathrm{R}}(\mathcal{C}) - 1)/2 \rfloor$, then $\boldsymbol{e}$ will be the unique element of rank at most

$r = \lfloor (d^{\mathrm{R}}_{\min}(\mathcal{C}) - 1)/2 \rfloor$ in $\mathcal{C}' = \langle \mathcal{C}, \boldsymbol{y} \rangle_{\mathbb{F}_q}$. Any element of $\mathcal{C}'$ has the form

$$M = x_0[\boldsymbol{y}] + \sum_{i=1}^{t} x_i M_i \in \mathbb{F}_q[x_0, \ldots, x_t]^{m \times n}.$$

Solving the search version of the MinRank problem with bound r is then equivalent to minimum distance decoding in $\mathcal{C}$.

The case when the underlying code is linear over the extension field, $\mathbb{F}_{q^m}$, is clearly a subset of the general MinRank problem. Any code of dimension k over $\mathbb{F}_{q^m}$ is a km-dimensional code over $\mathbb{F}_q$, and can be solved using any algorithm for solving the general MinRank problem. However, it is not known whether this subset of MinRank is **NP**-complete or not. The RSD problem can stated as follows.

Definition 5.1.1. Let $H \in \mathbb{F}_{q^m}^{(n-k)\times n}$, $\boldsymbol{s} \in \mathbb{F}_{q^m}^{n-k}$, and r a positive integer. The rank syndrome decoding (RSD) problem, $\mathcal{R}_{q^m,n,k}(H, \boldsymbol{s}, r)$, is to determine if there exists a $\boldsymbol{c} \in \mathbb{F}_{q^m}^{n}$ such that $\mathrm{rk}(\boldsymbol{c}) = r$ and $H\boldsymbol{c}^t = \boldsymbol{s}$.

We are only concerned with the case when r is the error correction capability of the code $\mathcal{C} = \langle H \rangle^{\perp}_{\mathbb{F}_{q^m}}$, in which case it is clear that $\mathcal{R}_{q^m,n,k}(H, \boldsymbol{s}, r)$ is equivalent to minimum distance decoding in $\mathcal{C}$. Chabaud and Stern showed in [12] that for an arbitrary H, $\mathcal{R}_{q^m,n,k}(H, \boldsymbol{s}, r)$ has complexity at most $O(q^{(m-r)(r-1)})$. Ourivski and Johansson improved this to $O(q^{(k+1)(r-1)})$ [61]. Gaborit et al. studied this problem in [30] and developed an algorithm which is exponential in complexity in general, but surprisingly, polynomial in complexity for a large range of parameters. His approach is in some ways a generalization of attacks based on guessing the error support, albeit adapted for the rank metric.

The first approach taken in [30] is to consider trapping the vector space support. Let $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e} \in \mathbb{F}_{q^m}^{n}$ be such that $\boldsymbol{x} \in \mathcal{C}$ and $\mathrm{wt}_{\mathrm{R}}(\boldsymbol{e}) = r$. If $E' \subseteq \mathbb{F}_{q^m}$ is a subspace such that $\mathrm{supp}_{\mathrm{vs}}(\boldsymbol{e}) \subseteq E'$, then we can reconstruct the coordinates of $\boldsymbol{e}$ from $\mathbb{F}_q$-combinations of elements of E'. Suppose that $\dim(E') = r' \geq r$ and E' contains each coordinate, e_i, of $\boldsymbol{e}$. if $\{E'_1, ..., E'_{r'}\}$ is a basis for E', then we can write

$$e_i = \sum_{j=1}^{r'} e_{ij} E'_j.$$

Then, $H\boldsymbol{y}^T = H\boldsymbol{e}^T$ and we obtain $n-k$ equations over $\mathbb{F}_{q^m}$, or $(n-k)m$ equations over $\mathbb{F}_q$. The number of unknowns (over $\mathbb{F}_q$) is $r'n$ (these are the $e_{i,j}$), and therefore we can solve this linear system if $r'n \leq (n-k)m$.

Since we do not know $\mathrm{supp}_{\mathrm{vs}}(\boldsymbol{e})$ a priori, one must enumerate and search through all possible E' in the hopes of finding one containing $\mathrm{supp}_{\mathrm{vs}}(\boldsymbol{e})$. The probability that $\mathrm{supp}_{\mathrm{vs}}(\boldsymbol{e})$ will be contained in a random r' dimensional subspace E' is

$$\frac{\begin{bmatrix} r' \\ r \end{bmatrix}_q}{\begin{bmatrix} m \\ r' \end{bmatrix}_q} \sim q^{-r(m-r')}.$$

Setting $r' = \left\lfloor \frac{(n-k)m}{n} \right\rfloor$ and considering also the matrix inversion, we obtain

$$\mathcal{R}_{q^m,n,k}(H, \boldsymbol{s}, r) \sim O\left((n-k)^3 m^3 q^{r\left\lceil \frac{km}{n} \right\rceil}\right),$$

for any $r \leq \dfrac{(n-k)m}{n}$.

An alternative attack, also proposed in [30] is to solve for the annulator polynomial of $\mathrm{supp}_{\mathrm{vs}}(\boldsymbol{e})$. Let $G \in \mathbb{F}_{q^m}^{k\times n}$, a generator matrix for $\langle H\rangle_{\mathbb{F}_{q^m}}^{\perp}$. Let $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}$ be the received vector, with $\boldsymbol{x} \in \langle G\rangle_{\mathbb{F}_{q^m}}$. If we denote by G_i the ith row of G, we can write

$$\boldsymbol{x} = \boldsymbol{m}G = \sum_{i=1}^{k} m_i G_i,$$

for some $\boldsymbol{m}$. Since $\boldsymbol{e}$ has rank weight r, there exists a monic linearized polynomial $P = P_0 x + P_1 x^{[1]} + ... + P_{r-1}x^{[r-1]} + x^{[r]}$ of degree $[r]$ vanishing on $\langle e_1, \ldots, e_n\rangle_{\mathbb{F}_q}$. This gives,

$$P\left(y_j - \sum_{i=1}^{k} m_i g_{ij}\right) = 0. \tag{5.1}$$

From here we obtain a system of n equations with $k + r$ unknowns, the message coordinates m_i, and the coefficients of P. The equations we obtain are not linear, but they are sparse. Overall, there are $(k + 1)(r + 1) - 1$ terms. After linearizing, one can obtain a unique solution to this system if it is full rank. In [30], it is observed that (5.1) appears to be generically full rank, and therefore we obtain following result.

Proposition 5.1.2 ([30])**.** *Let $\mathcal{C}$ be a code with parity check matrix $H \in \mathbb{F}_{q^m}^{(n-k)\times n}$ capable of correcting any error pattern, $\boldsymbol{e}$, of rank weight r. Let $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}$ where $\boldsymbol{x} \in \mathcal{C}$. If k satisfies $(r + 1)(k + 1) - 1 \leq n$, and system (5.1) is a full rank linear system, then*

$$\mathcal{R}_{q^m,n,k}(H, H\boldsymbol{y}^T, r) \sim O(((r + 1)(k + 1) - 1)^3).$$

An extension of Proposition 5.1.2 by guessing errors allowed the authors of [30] to obtain the following corollary which we will call the *error support (ES)* attack.

Corollary 5.1.3. *Let $\mathcal{C} \subset \mathbb{F}_{q^m}^n$ be a code of dimension k capable of correcting r errors with parity check matrix, H, and such that any generator matrix for $\mathcal{C}$ has column rank n. If*

$$\left\lceil \frac{(r + 1)(k + 1) - (n + 1)}{r} \right\rceil \leq k,$$

then

$$\mathcal{R}_{q^m,n,k}(H, \boldsymbol{s}, r) \sim O(r^3 k^3 q^{r\left\lceil \frac{(r+1)(k+1)-(n+1)}{r} \right\rceil}).$$

Other generic attacks, for instance [61], are based on the idea of guessing the Grassmann support, rather than the vector space support of an element.

5.2 Preliminaries

Recall from Corollary 3.1.13 that a matrix $M \in \mathbb{F}_{q^m}^{k\times n}$ of rank t can be expressed as $M = M'U$, for $M' \in \mathbb{F}_{q^m}^{k\times t}$ and $U \in \mathbb{F}_q^{t\times n}$, and $\langle U\rangle_{\mathbb{F}_{q^m}}$ is called the Grassmann support of M. We will also refer to U, the matrix, as a Grassmann support matrix. Of course, what we mean is that U is a matrix representation for the Grassmann support of M with entries in $\mathbb{F}_q$. We begin with the following important observation.

Lemma 5.2.1. *Let $G \in \mathbb{F}_{q^m}^{k\times n}$ be an arbitrary generator matrix for a Gabidulin code, not necessarily in the form of (3.9). Then, we can recover a polynomial time decoding algorithm with respect to G.*

Proof. We first note from if the generating element of a Gabidulin code is known, then an efficient decoding algorithm can be given for G of the form (3.9) [77]. Suppose that $\langle G\rangle_{\mathbb{F}_{q^m}}$ is the Gabidulin code $\mathrm{Gab}_{n,k}(\boldsymbol{\alpha})$ with dimension $1 < k < n$ and generator matrix $S\overline{G}$, where $S \in \mathrm{GL}_k(\mathbb{F}_{q^m})$ and $\overline{G}$ of the form (3.9). Note that $\mathrm{Gab}_{n,k}(\boldsymbol{\alpha})^{([1])} \cap \mathrm{Gab}_{n,k}(\boldsymbol{\alpha})$ is the Gabidulin code $\mathrm{Gab}_{n,k-1}(\boldsymbol{\alpha}^{([1])})$ (see Lemma 3.2.10). Iterating with this new Gabidulin code, we can eventually obtain a code of dimension 1, which is generated by $\boldsymbol{\alpha}^{([k-1])}$. If we take some non-zero element of this space, it has the form $\beta\boldsymbol{\alpha}^{([k-1])}$, for some $\beta \in \mathbb{F}_{q^m}$. Applying the Frobenius map coordinate-wise $m-k+1$ times, we obtain an element of the form $\beta^{[m-k+1]}\boldsymbol{\alpha}$. Using this element, we can construct a generator matrix, for $\mathrm{Gab}_{n,k}(\boldsymbol{\alpha})$ having the form

$$\underbrace{\begin{pmatrix} \beta^{[m-k+1]} & & & \\ & \beta^{[m-k+2]} & & \\ & & \ddots & \\ & & & \beta \end{pmatrix}}_{B} \underbrace{\begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \alpha_1^{[1]} & \alpha_2^{[1]} & \dots & \alpha_n^{[1]} \\ & & \vdots & \\ \alpha_1^{[k-1]} & \alpha_2^{[k-1]} & \dots & \alpha_n^{[k-1]} \end{pmatrix}}_{\overline{G}}.$$

The change of basis from $S\overline{G}$ to $B\overline{G}$ is then given by BS^{-1}. For a message $\boldsymbol{m} \in \mathbb{F}_{q^m}^k$, encoded as $\boldsymbol{m}S\overline{G}$, we can now decode with respect to $\mathrm{Gab}_{n,k}(\beta^{[m-k+1]}\boldsymbol{\alpha})$ to obtain $\boldsymbol{m}SB^{-1}$. Then, applying BS^{-1}, we can recover $\boldsymbol{m}$. □

Lemma 5.2.2. *Let $X \in \mathbb{F}_{q^m}^{k\times n}$ be a matrix of rank k and column rank $t \geq k$ with Grassmann support matrix $U \in \mathbb{F}_q^{t\times n}$. Then $\langle X\rangle_{\mathbb{F}_{q^m}} \subseteq \langle U\rangle_{\mathbb{F}_{q^m}}$ and the inclusion is strict if and only if $t > k$.*

Proof. By Corollary 3.1.13, we can write $X = VU$ for some $V \in \mathbb{F}_{q^m}^{k\times t}$. Thus, every row of X is a $\mathbb{F}_{q^m}$-linear combination of the rows of U, which implies that $\langle X\rangle_{\mathbb{F}_{q^m}} \subseteq \langle U\rangle_{\mathbb{F}_{q^m}}$. Since $\dim\left(\langle U\rangle_{\mathbb{F}_{q^m}}\right) = t$ and $\dim\left(\langle X\rangle_{\mathbb{F}_{q^m}}\right) = k$, we get equality if and only if $k = t$. □

Lemma 5.2.3. *Let $X \in \mathbb{F}_{q^m}^{k\times n}$ be a matrix of column rank t and $S \in \mathrm{GL}_k(\mathbb{F}_{q^m})$. Then, SX also has column rank t.*

Proof. Denote the ith column of X by X_i. Assume that SX has column rank less than t, i.e. for any $i_1 < \cdots < i_t \in \{1, \ldots, n\}$ there exists $a_1, \ldots, a_t \in \mathbb{F}_q$ such that

$$\sum_{\ell=1}^{t} a_\ell (SX)_{i_\ell} = 0 \Leftrightarrow S \sum_{\ell=1}^{t} a_\ell X_{i_\ell} = 0 \Leftrightarrow \sum_{\ell=1}^{t} a_\ell X_{i_\ell} = 0.$$

This is a contradiction to the fact that the column rank of X is t, therefore SX also must have column rank t. □

We saw that if a matrix X, with Grassmann support matrix U, has column rank which is greater than its rank, then $\langle X \rangle_{\mathbb{F}_{q^m}} \subsetneq \langle U \rangle_{\mathbb{F}_{q^m}}$. The following theorem shows that we can use the Frobenius map to recover $\langle U \rangle_{\mathbb{F}_{q^m}}$ from X.

Theorem 5.2.4. *Let $X \in \mathbb{F}_{q^m}^{k \times n}$ be a matrix of column rank s with Grassmann support matrix $U \in \mathbb{F}_q^{s \times n}$ and let $s' \geq s$. Then,*

$$\sum_{i=0}^{s'-1} \langle X \rangle_{\mathbb{F}_{q^m}}^{([i])} = \langle U \rangle_{\mathbb{F}_{q^m}}.$$

In particular,

$$\dim \left(\sum_{i=0}^{s'-1} \langle X \rangle_{\mathbb{F}_{q^m}}^{([i])} \right) = s.$$

Proof. We prove it for $s' = s$. For the case $s' > s$, successive coordinate-wise Frobenius powers fix the space and therefore the statement remains true. Consider

$$\langle X \rangle_{\mathbb{F}_{q^m}} \subseteq \langle X \rangle_{\mathbb{F}_{q^m}} + \langle X \rangle_{\mathbb{F}_{q^m}}^{([1])} \subseteq \cdots \subseteq \sum_{i=0}^{s-1} \langle X \rangle_{\mathbb{F}_{q^m}}^{([i])} \subseteq \langle U \rangle_{\mathbb{F}_{q^m}}.$$

By considering the dimension of each of the spaces in the chain above, we see that there must exist $\ell \leq s$ be such that,

$$\sum_{i=0}^{\ell-1} \langle X \rangle_{\mathbb{F}_{q^m}}^{([i])} = \sum_{i=0}^{\ell} \langle X \rangle_{\mathbb{F}_{q^m}}^{([i])}.$$

Define $s' = \dim \left(\sum_{i=0}^{\ell-1} \langle X \rangle_{\mathbb{F}_{q^m}}^{([i])} \right)$. We note that $s' \leq s$ since, for each i, $\langle X \rangle_{\mathbb{F}_{q^m}}^{([i])} \subset \langle U \rangle_{\mathbb{F}_{q^m}}$ which has dimension s. We have

$$\left(\sum_{i=0}^{\ell-1} \langle X \rangle_{\mathbb{F}_{q^m}}^{([i])} \right)^{([1])} = \sum_{i=1}^{\ell} \langle X \rangle_{\mathbb{F}_{q^m}}^{([i])} \subseteq \sum_{i=0}^{\ell} \langle X \rangle_{\mathbb{F}_{q^m}}^{([i])} = \sum_{i=0}^{\ell-1} \langle X \rangle_{\mathbb{F}_{q^m}}^{([i])}. \tag{5.2}$$

Since the coordinate-wise Frobenius map preserves the dimension of spaces we must have equality in (5.2). By Proposition 3.1.12 and the third point of Lemma 3.1.8, we can express the sum on the right as the row space of a matrix $U' \in \mathbb{F}_q^{s' \times n}$ of (column) rank s'. This implies that $\langle X \rangle_{\mathbb{F}_{q^m}} \subseteq \langle U' \rangle_{\mathbb{F}_{q^m}}$. It follows from Proposition 3.1.12 that $s' \geq s$. Therefore, we must have $s = s'$ and therefore $\langle U \rangle_{\mathbb{F}_{q^m}} = \langle U' \rangle_{\mathbb{F}_{q^m}}$. Since $\ell \leq s$, we obtain the result. □

Theorem 5.2.4 simply says that if one takes enough coordinate-wise Frobenius images of a matrix, then they generate the Grassmann support. Now, we note that a matrix $X \in \mathbb{F}_{q^m}^n$ can always be decomposed into a Moore matrix component X_{Moore} and a non-Moore matrix component Z as

$$X = X_{\text{Moore}} + Z.$$

We will call such a decomposition a *Moore decomposition*.

Definition 5.2.5. A *minimum column rank Moore decomposition* is a Moore decomposition in which the non-Moore component has lowest possible column rank.

Proposition 5.2.6 shows that, regardless of the choice of Moore decomposition, the Grassmann support of a non-Moore matrix component of a minimum column rank Moore decomposition is invariant with respect the decomposition.

Proposition 5.2.6. *Suppose that $X \in \mathbb{F}_{q^m}^{k\times n}$ is a matrix which has minimum column rank Moore decomposition $X = A_{\text{Moore}}+A$, where A_{Moore} is a Moore matrix, and A has column rank s. Then, any other minimum column rank Moore decomposition, $X = B_{\text{Moore}} + B$, satisfies that* $\text{supp}_{\text{Gr}}(A) = \text{supp}_{\text{Gr}}(B)$.

Proof. Let A have Grassmann support matrix U, and B have Grassmann support matrix V. Write $A = A'U$ and $B = B'V$ with $U, V \in \mathbb{F}_q^{s\times n}$ according to Proposition 3.1.12. Let $E \in \mathbb{F}_q^{(n-s)\times n}$ be a parity check matrix for $\langle V\rangle_{\mathbb{F}_{q^m}}$. Then,

$$B_{\text{Moore}}E^T = XE^T - BE^T = XE^T = A_{\text{Moore}}E^T + AE^T,$$

which yields

$$(B_{\text{Moore}} - A_{\text{Moore}})E^T = AE^T.$$

Since E is a matrix over $\mathbb{F}_q$, $(B_{\text{Moore}} - A_{\text{Moore}})E^T$ is a Moore matrix, therefore the matrix AE^T must be a Moore matrix as well. The ith row of AE^T can be written as $(AE^T)_i = (A_1E^T)^{([i-1])}$. Since A itself is not necessarily a Moore matrix, row i of A must be of the form $A_i \in A_1^{([i-1])} + \ker(E^T)$, for $i = 1, \ldots, k$. Then, we can write

$$A = \underbrace{\begin{pmatrix} A_1 \\ A_1^{([1])} \\ \vdots \\ A_1^{([k-1])} \end{pmatrix}}_{\bar{A}} + \underbrace{\begin{pmatrix} \kappa_1 \\ \kappa_2 \\ \vdots \\ \kappa_k \end{pmatrix}}_{\kappa' V},$$

for $\kappa_i \in \ker(E^T) = \langle V\rangle_{\mathbb{F}_{q^m}}$. If we let $F \in \mathbb{F}_q^{(n-s)\times n}$ be a parity check matrix for $\langle U\rangle_{\mathbb{F}_{q^m}}$, then $AF^T = 0$ and hence $A_1F^T = 0$. Since F is a matrix over $\mathbb{F}_q$, we also get $A_1^{([i])}F^T = 0$ for $i = 1, \ldots, k-1$. Hence,

$$0 = AF^T = \bar{A}F^T + \kappa' VF^T = \kappa' VF^T. \tag{5.3}$$

Since $X = (A_{\text{Moore}} + \bar{A}) + \kappa' V$ is also a Moore decomposition of X, then the column rank of $\kappa' V$ must at least s and since $\langle V \rangle_{\mathbb{F}_{q^m}} \supseteq \langle \kappa' V \rangle_{\mathbb{F}_{q^m}}$ we see that they must actually be equal. Therefore, from Equation (5.3),

$$\langle V \rangle_{\mathbb{F}_{q^m}} F^T = 0,$$

and therefore, $\langle V \rangle_{\mathbb{F}_{q^m}} = \langle U \rangle_{\mathbb{F}_{q^m}}$, so the Grassmann supports of A and B are the same. □

Theorem 5.2.7. *Let $M \in \mathbb{F}_{q^m}^{k \times n}$ be a Moore matrix and $X \in \mathbb{F}_{q^m}^{k \times n}$ be of column rank s, where s is the rank of the non-Moore component in a minimum column rank Moore decomposition of X. Let $U \in \mathbb{F}_q^{s \times n}$ be a Grassmann support matrix for X. Then, we have*

$$\sum_{i=0}^{s} \langle M + X \rangle_{\mathbb{F}_{q^m}}^{([i])} = \langle U \rangle_{\mathbb{F}_{q^m}} + \sum_{i=0}^{s} \langle M \rangle_{\mathbb{F}_{q^m}}^{([i])}.$$

Proof. Let X_i, M_i denote the ith row of X and M respectively, and let

$$X' = \begin{pmatrix} X_1^{([1])} - X_2 \\ X_2^{([1])} - X_3 \\ \vdots \\ X_{k-1}^{([1])} - X_k \end{pmatrix}, M^* = \begin{pmatrix} M_1 \\ M_1^{([1])} \\ \vdots \\ M_1^{([k+s-1])} \end{pmatrix}, X^* = \left(\begin{array}{c} X_1 \\ X_1^{([1])} \\ \vdots \\ X_1^{([s-1])} \\ \hline X_1^{([s])} \\ X_2^{([s])} \\ \vdots \\ X_k^{([s])} \end{array}\right).$$

Then, the space $\sum_{i=0}^{s} \langle M + X \rangle_{\mathbb{F}_{q^m}}^{([i])}$ is generated by the row span of

$$\begin{pmatrix} M + X \\ (M + X)^{([1])} \\ \vdots \\ (M + X)^{([s])} \end{pmatrix} = \tilde{S} \begin{pmatrix} M^* + X^* \\ X' \\ \vdots \\ (X')^{([s-1])} \end{pmatrix}, \tag{5.4}$$

for a suitable row transformation matrix $\tilde{S}$. We note that each row of X' must be contained in $\langle U \rangle_{\mathbb{F}_{q^m}}$, and therefore, each element of $(X')^{([i])}$ must be as well. If we let U' be a Grassmann support matrix of X', we must have $\dim\left(\langle U' \rangle_{\mathbb{F}_{q^m}}\right) = s' \leq s$. By Theorem 5.2.4,

$$\sum_{i=0}^{s'-1} \langle X' \rangle_{\mathbb{F}_{q^m}}^{([i])} = \sum_{i=0}^{s-1} \langle X' \rangle_{\mathbb{F}_{q^m}}^{([i])} = \langle U' \rangle_{\mathbb{F}_{q^m}} \subseteq \langle U \rangle_{\mathbb{F}_{q^m}}.$$

We now want to show that $\langle U' \rangle_{\mathbb{F}_{q^m}} = \langle U \rangle_{\mathbb{F}_{q^m}}$. Suppose for the sake of contradiction that the rank of U' is strictly smaller than s. We write X as a Moore decomposition

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix} = \begin{pmatrix} X_1 \\ X_1^{([1])} \\ \vdots \\ X_1^{([k-1])} \end{pmatrix} + \underbrace{\begin{pmatrix} 0 \\ X_2 - X_1^{([1])} \\ \vdots \\ X_k - X_1^{([k-1])} \end{pmatrix}}_{X''}.$$

We note that $X_{i+1} - X_i^{[1]} \in \langle U' \rangle_{\mathbb{F}_{q^m}}$ for $i = 1, \ldots, k-1$. Starting from the first non-zero row of X'', it follows that

$$(X_2 - X_1^{([1])})^{([1])} = X_2^{([1])} - X_1^{([2])} \in \langle U' \rangle_{\mathbb{F}_{q^m}}$$

which implies

$$\begin{aligned} & X_2^{([1])} - X_1^{([2])} - (X_2^{([1])} - X_3) \in \langle U' \rangle_{\mathbb{F}_{q^m}} \\ \Leftrightarrow \quad & X_3 - X_1^{([2])} \in \langle U' \rangle_{\mathbb{F}_{q^m}}. \end{aligned}$$

We recognize this as the second non-zero row of X''. Continuing in this fashion, we can obtain that every row of X'' must belong to $\langle U' \rangle_{\mathbb{F}_{q^m}}$. Hence, X'' has column rank at most $s' < s$. However, this contradicts the fact that the minimal column rank Moore decomposition has non-Moore component with column rank s. Therefore, by Proposition 5.2.6, U' has rank s and we have $\langle U' \rangle_{\mathbb{F}_{q^m}} = \langle U \rangle_{\mathbb{F}_{q^m}}$.

Hence, the row space of the second matrix in (5.4) is equal to the row space of

$$\begin{pmatrix} M^* + X^* \\ U \end{pmatrix}$$

which is in turn equal to the row space of

$$\begin{pmatrix} M^* \\ U \end{pmatrix},$$

because we can cancel X^* by taking suitable elements of $\langle U \rangle_{\mathbb{F}_{q^m}}$, since $\langle X^* \rangle_{\mathbb{F}_{q^m}} \subseteq \langle U' \rangle_{\mathbb{F}_{q^m}} = \langle U \rangle_{\mathbb{F}_{q^m}}$. The result now follows. □

Lemma 5.2.8. *Let $X \in \mathbb{F}_{q^m}^{k \times n}$ have minimum column rank Moore decomposition, $X = X_{\text{Moore}} + Z$. Then,*

$$\text{supp}_{\text{Gr}}(Z) + \text{supp}_{\text{Gr}}(X_{\text{Moore}}) = \text{supp}_{\text{Gr}}(X).$$

Proof. Denote by $\ell = \max\{\text{colrk}_{\mathbb{F}_q}(X), \text{colrk}_{\mathbb{F}_q}(X_{\text{Moore}}), \text{colrk}_{\mathbb{F}_q}(Z)\}$. Using Theorems 5.2.4 and 5.2.7 we have,

$$\begin{aligned}
\text{supp}_{\text{Gr}}(X) &= \sum_{i=0}^{\ell} \langle X \rangle_{\mathbb{F}_{q^m}}^{([i])} \\
&= \sum_{i=0}^{\ell} \langle X_{\text{Moore}} + Z \rangle_{\mathbb{F}_{q^m}}^{([i])} \\
&= \sum_{i=0}^{\ell} \langle X_{\text{Moore}} \rangle_{\mathbb{F}_{q^m}} + \text{supp}_{\text{Gr}}(Z). \\
&= \text{supp}_{\text{Gr}}(X_{\text{Moore}}) + \text{supp}_{\text{Gr}}(Z).
\end{aligned}$$

□

Corollary 5.2.9. *Let $X \in \mathbb{F}_{q^m}^{k \times n}$ have minimum column rank Moore decomposition, $X = X_{\text{Moore}} + Z$. Then,*

$$\text{colrk}_{\mathbb{F}_q}(Z) \leq \text{colrk}_{\mathbb{F}_q}(X).$$

Corollary 5.2.10. *Let $M \in \mathbb{F}_{q^m}^{k \times N}$ be a Moore matrix and X be of column rank t with minimum column rank Moore decomposition $X = X_{\text{Moore}} + Z$, where $\text{colrk}_{\mathbb{F}_q}(Z) = s$. Suppose that $d_{\min}^{\text{R}}(\langle M \rangle_{\mathbb{F}_{q^m}}) \geq s + t + 2$. Let $\mathcal{U}$ denote the space spanned by the elements of rank one in*

$$\sum_{i=0}^{s} \langle M + X \rangle_{\mathbb{F}_{q^m}}^{([i])}.$$

Then we have,

$$\text{supp}_{\text{Gr}}(Z) \subseteq \mathcal{U} \subseteq \text{supp}_{\text{Gr}}(X).$$

Moreover, if $s = t$, then $\text{supp}_{\text{Gr}}(Z) = \mathcal{U} = \text{supp}_{\text{Gr}}(X)$.

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

$$\sum_{i=0}^{s} \langle M + X \rangle_{\mathbb{F}_{q^m}}^{([i])}.$$

From Lemma 5.2.8, if $X = X_{\text{Moore}} + Z$, is a minimum column rank decomposition, then we know that $\text{supp}_{\text{Gr}}(Z) \subseteq \text{supp}_{\text{Gr}}(X)$. Let $H \in \mathbb{F}_q^{(n-t) \times n}$ be parity check matrix for $\text{supp}_{\text{Gr}}(X)$. Then, we have

$$\begin{aligned}
d_{\min}^{\text{R}} \left(\sum_{i=0}^{s} \langle M + X \rangle_{\mathbb{F}_{q^m}}^{([i])} H^T \right) &= d_{\min}^{\text{R}} \left(\sum_{i=0}^{s} \langle M \rangle_{\mathbb{F}_{q^m}}^{([i])} H^T \right) \\
&\geq (s + t + 2) - s - t \\
&= 2,
\end{aligned}$$

where the inequality comes from Lemma 3.2.10. Since $\mathrm{wt_R}(\boldsymbol{x}) \geq \mathrm{wt_R}(\boldsymbol{x}H^T)$, we must have that all elements of $\mathcal{U}$ must belong to $\mathrm{supp_{Gr}}(X)$. By Theorem 5.2.7, $\mathrm{supp_{Gr}}(Z) \subset \mathcal{U}$ and therefore, if $s = t$ they are equal. □

Finally, we need the following lemma, which allows us to efficiently compute the elements of rank one in a linear rank metric code. To accomplish this, we only need to find the codewords that have all coordinates in $\mathbb{F}_q$ (all other rank one codewords are $\mathbb{F}_{q^m}$-multiples of these).

Lemma 5.2.11. *Let $G \in \mathbb{F}_{q^m}^{k\times n}$ be in reduced row echelon form and denote by G_i the ith row of G. Then the solutions to*

$$\sum_{i=1}^{k} a_i(G_i^{([1])} - G_i) = 0, \tag{5.5}$$

for variables $a_i \in \mathbb{F}_q$, represent the codewords of $\langle G\rangle_{\mathbb{F}_{q^m}} \cap \mathbb{F}_q^n$. Hence, a basis for the space spanned by the elements of rank one in G can be found in polynomial time requiring $O(kmn^2)$ operations over $\mathbb{F}_q$.

Proof. Any codeword can be written as an $\mathbb{F}_{q^m}$-linear combination of the rows of G. Since all rows of G have their pivot equal to 1, a codeword with entries only in $\mathbb{F}_q$ needs to be an $\mathbb{F}_q$-linear combination of the rows. Thus, we get that any codeword in $\mathbb{F}_q^n$ can be written as $\sum_{i=1}^{k} a_iG_i$ for some $a_i \in \mathbb{F}_q$. Furthermore, we know that

$$\boldsymbol{v} \in \mathbb{F}_q^n \Leftrightarrow \boldsymbol{v}^{([1])} - \boldsymbol{v} = 0,$$

hence

$$\sum_{i=1}^{k} a_iG_i \in \mathbb{F}_q^n \Leftrightarrow \sum_{i=1}^{k} a_i(G_i^{([1])} - G_i) = 0.$$

In order to solve Equation (5.5), we require kn Frobenius maps in $\mathbb{F}_{q^m}$ and a matrix reduction on a matrix over $\mathbb{F}_q$ of size $km \times n$ and so finding the elements requires on the order of kmn^2 operations over $\mathbb{F}_q$. □

An algorithm for computing the elements of rank one from a matrix G is given in Appendix A.1. We note that this algorithm only works for rank metric codes which are linear over $\mathbb{F}_{q^m}$. For rank metric codes which are linear over $\mathbb{F}_q$ and not over $\mathbb{F}_{q^m}$, the space spanned by an element of rank one may not include a vector belonging to $\mathbb{F}_q^n$. Therefore, for a code which is linear over $\mathbb{F}_q$, we could repeat the algorithm, but we would have to solve for all $\mathbb{F}_q$-lines, $\ell_\alpha = \{\alpha \cdot a \mid a \in \mathbb{F}_q\}$, which corresponds to solving $x^{[1]} - \alpha^{[1]-1}x$ for $(q^m - 1)/(q - 1)$ values of α.

5.3 Cryptanalysis of GPT Cryptosystem

We note that for most parameters, the GPT cryptosystem was effectively broken by Overbeck in [62]. Nevertheless, it is useful to demonstrate the effectiveness of our attack first on the GPT cryptosystem before moving on to attack the variants which do resist Overbeck's attack. We note that our attack is still more general–we can break the GPT cryptosystem for all parameters.

Recall that the public key generator matrix is of the form

$$G_{\text{pub}} = SG + X \in \mathbb{F}_{q^m}^{k\times n},$$

where G is a generator matrix of some Gabidulin code $\text{Gab}_{n,k}(\boldsymbol{\alpha})$ capable of correcting t' errors, $X \in \mathbb{F}_{q^m}^{k\times n}$ is a matrix of column rank $t < t'$, and $S \in \text{GL}_k(\mathbb{F}_{q^m})$. Let $X = X_{\text{Moore}} + Z$ be a minimum column rank Moore decomposition with $\text{colrk}_{\mathbb{F}_q}(Z) = s$.

Note that, as an attacker, we do not have a priori knowledge of the parameter s (the column rank of the non-Moore part in the minimal column rank Moore decomposition of X). We can generally assume $s = t$, or else start with $s = t$ and decrease the value until the attack succeeds.

Theorem 5.3.1. *Consider a GPT cryptosystem as defined in Subsection 4.2.1, where $S^{-1}X = X_{\text{Moore}} + Z$ is a minimal column rank Moore decomposition. Suppose an adversary can find a full rank matrix $U' \in \mathbb{F}_q^{s'\times n}$ satisfying*

$$\text{supp}_{\text{Gr}}(Z) \subseteq \langle U'\rangle_{\mathbb{F}_{q^m}} \subseteq \text{supp}_{\text{Gr}}(X),$$

then an encrypted message from a public key of the form (4.2) *can be recovered in polynomial time.*

Proof. Let $H \in \mathbb{F}_q^{(n-s')\times n}$ be a parity check matrix for $\langle U'\rangle_{\mathbb{F}_{q^m}}$. Applying H^T to the public key generator matrix yields

$$G_{\text{pub}}H^T = (SG + X)H^T = S(G + X_{\text{Moore}})H^T.$$

From Lemma 5.2.8 we know that $\text{colrk}(X_{\text{Moore}}) \leq t$. Then, from Lemma 3.2.10, it follows that $\langle G + X_{\text{Moore}}\rangle_{\mathbb{F}_{q^m}}$ has minimum rank distance at least $n-k+1-t$, and that $\langle G + X_{\text{Moore}}H^T\rangle_{\mathbb{F}_{q^m}}$ has minimum rank distance at least $n-k+1-(t+s')$. Moreover, $GH^T + X_{\text{Moore}}H^T$ is a Moore matrix.

From the minimum distance we know that there are $n-(t+s')$ independent columns in this matrix, which generate a Gabidulin code of minimum distance $n-(t+s')-k+1$, $\text{Gab}_{n-(t+s'),k}(\boldsymbol{\gamma})$, for some $\boldsymbol{\gamma} \in \mathbb{F}_{q^m}^{n-(t+s')}$. From Lemma 5.2.1, we can recover a decoding algorithm for $\text{Gab}_{n-(t+s'),k}(\boldsymbol{\gamma})$ with respect to the submatrix formed by these $n-(t+s')$ columns. The error correction capability of $\text{Gab}_{n-(t+s'),k}(\boldsymbol{\gamma})$ is

$$\left\lfloor \frac{n-(t+s')-k}{2} \right\rfloor = \left\lfloor t' - \frac{t+s'}{2} \right\rfloor \geq t' - t \geq \text{rk}(\boldsymbol{e}) \geq \text{rk}(\boldsymbol{e}H^T),$$

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

$$(\boldsymbol{m}(SG+X)+\boldsymbol{e})H^T = \boldsymbol{m}S(GH^T + X_{\text{Moore}}H^T) + \boldsymbol{e}H^T.$$

When we restrict this to the above chosen independent columns, we can decode with respect to the the submatrix generating $\text{Gab}_{n-(t+s'),k}(\boldsymbol{\gamma})$ and can therefore recover $\boldsymbol{m}$. All operations above are polynomial. □

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

Corollary 5.3.2. *Consider a GPT cryptosystem as defined in Subsection 4.2.1 with public key generator matrix $G_{\text{pub}} = SG + X \in \mathbb{F}_{q^m}^{k\times n}$. For any such cryptosystem, an encrypted message can be recovered in polynomial time.*

Proof. Let $S^{-1}X = X_{\text{Moore}} + Z$ be a minimal column rank Moore decomposition. Denote by s the column rank of Z. We first note that $d^{\text{R}}_{\min}(\langle G\rangle_{\mathbb{F}_{q^m}}) \geq s+t+2$ is always satisfied. To see this, we have

$$\frac{d^{\text{R}}_{\min}(\langle G\rangle_{\mathbb{F}_{q^m}})-1}{2} \geq \left\lfloor \frac{n-k}{2} \right\rfloor = t' > t \geq \frac{s+t}{2}.$$

By Corollary 5.2.10, all the elements of rank one in

$$\sum_{i=0}^{s} \langle G + X_{\text{Moore}} + Z\rangle^{([i])}_{\mathbb{F}_{q^m}} = \sum_{i=0}^{s} \langle SG+X\rangle^{([i])}_{\mathbb{F}_{q^m}}$$

belong to the Grassmann support of X, and therefore from Lemma 5.2.11 we can find a matrix U' with entries in $\mathbb{F}_q$, the rows of which are a basis for the elements of rank one in $\sum_{i=0}^{s}\langle SG+X\rangle^{([i])}_{\mathbb{F}_{q^m}}$. Furthermore, $\dim(\langle U'\rangle_{\mathbb{F}_{q^m}}) = s' \leq t$ and $\text{supp}_{\text{Gr}}(Z) \subset \langle U'\rangle_{\mathbb{F}_{q^m}}$ from Theorem 5.2.7. Using Theorem 5.3.1 we can recover the message. □

5.4 Cryptanalysis of GGPT Variants

In this section, we show how one can use information about the elements of rank one in order to recover some structure about the public key. As the public key contains no elements of rank one by construction, some transformation of the code must be performed which predictably changes the structure so that the elements of rank one leak some information which we can exploit to determine the structure of the system.

5.4.1 LGGPT Variant Cryptanalysis

We first consider the LGGPT variant. Recall that the LGGPT variant has a public key of the form

$$\hat{G}_{\text{pub}} = S[X \mid G]\sigma,$$

where $G \in \mathbb{F}_{q^m}^{k\times n}$ is a generator matrix of a Gabidulin code, $S \in \mathrm{GL}_k(\mathbb{F}_{q^m})$, $\sigma \in \mathrm{GL}_n(\mathbb{F}_q)$, and X is a randomly chosen matrix of column rank $\hat{t}$ over $\mathbb{F}_q$ and rank a. We saw in Section 4.2.1 that a must satisfy

$$a(n-k) < \hat{t}.$$

Under this constraint, we can not recover the structure of X without interference from the structure of G. In particular, we have

$$\sum_{i=0}^{\lfloor \frac{\hat{t}}{a} \rfloor} \langle [X \mid 0]\sigma \rangle_{\mathbb{F}_{q^m}}^{([i])} \subsetneq \mathrm{supp}_{\mathrm{Gr}}([X \mid 0]\sigma),$$

but

$$\mathrm{supp}_{\mathrm{Gr}}([0 \mid G]\sigma) = \sum_{i=0}^{n-k} \langle [0 \mid G]\sigma \rangle_{\mathbb{F}_{q^m}}^{([i])} \subset \sum_{i=0}^{\lfloor \frac{\hat{t}}{a} \rfloor} \langle [0 \mid G]\sigma \rangle_{\mathbb{F}_{q^m}}^{([i])}.$$

Therefore, Overbeck's attack will certainly fail, since we cannot recover the structure of the designed error part before we obtain the structure of the Gabidulin part of the public key.

Our strategy will be to recover $\mathrm{supp}_{\mathrm{Gr}}([0 \mid G]\sigma)$ in some extended matrix. Before we can proceed, we will need to make some assumptions regarding the behavior of X as well as that of random subcodes of Gabidulin codes. First, note that there is a suitable row transformation, $T \in \mathrm{GL}_n(\mathbb{F}_{q^m})$, so that

$$T\hat{G}_{\mathrm{pub}} = \left(\begin{array}{c|c} X^* & G^* \\ 0 & G^{**} \end{array} \right) \sigma, \tag{5.6}$$

where $X^* \in \mathbb{F}_{q^m}^{a\times \hat{t}}$ is a matrix with the same row span as X, and G^* and G^{**} are matrices which span subcodes of $\langle G \rangle_{\mathbb{F}_{q^m}}$. One can easily see that $\langle [X^* \mid G^*]\sigma \rangle_{\mathbb{F}_{q^m}}$ and $\langle [0 \mid G^{**}]\sigma \rangle_{\mathbb{F}_{q^m}}$ intersect trivially and we can therefore ignore σ for the time being.

We want to be able to generate $\mathrm{supp}_{\mathrm{Gr}}([0 \mid G])$ from $\langle G^{**} \rangle_{\mathbb{F}_{q^m}}$. We note that if

$$G^{**} = \begin{pmatrix} \boldsymbol{g} \\ \vdots \\ \boldsymbol{g}^{([k-a-1])} \end{pmatrix},$$

then we can recover $\mathrm{supp}_{\mathrm{Gr}}([0 \mid G])$ as

$$\mathrm{supp}_{\mathrm{Gr}}([0 \mid G]) = \sum_{i=0}^{\lceil \frac{n}{k-a} \rceil - 1} \langle G^{**} \rangle_{\mathbb{F}_{q^m}}^{([i(k-a)])}.$$

Therefore we make the following assumption. Table 5.1 gives some evidence for why these assumptions appear to be generically satisfied.

Assumption 5.4.1. *Let $G \in \mathbb{F}_{q^m}^{k\times n}$ be a generator matrix of a Gabidulin code, and $\mathcal{B} \subset \langle G\rangle_{\mathbb{F}_{q^m}}$ be a random subspace of $\langle G\rangle_{\mathbb{F}_{q^m}}$ of codimension a. Set*

$$\ell = \left\lceil \frac{n}{k-a} \right\rceil. \tag{5.7}$$

With high probability, we have

$$\sum_{i=0}^{\ell-1} \mathcal{B}^{([i(k-a)])} = \mathbb{F}_{q^m}^n. \tag{5.8}$$

The value of ℓ in (5.7) is the smallest possible value for which we can obtain equality in (5.8). One could choose ℓ larger than in (5.7), although in doing so there is a trade off. This is related to our next assumption.

Assumption 5.4.2. *Let $X \in \mathbb{F}_{q^m}^{k\times \hat{t}}$ be a random matrix of rank a. For ℓ given in* (5.7)*, if $\ell a \ll \hat{t}$, then with high probability,*

$$\sum_{i=0}^{\ell-1} \langle X\rangle_{\mathbb{F}_{q^m}}^{([i(k-a)])}$$

contains no elements of rank one.

The idea is that since we expect a random space of high codimension not to contain elements of rank one, we want to keep

$$\sum_{i=0}^{\ell-1} \langle [X^* \mid G^*]\rangle_{\mathbb{F}_{q^m}}^{[i(k-a)]}$$

small whilst still being able to construct $\mathrm{supp}_{\mathrm{Gr}}([0 \mid G])$. This is summarized in the following theorem.

Theorem 5.4.3. *Let $S \in \mathrm{GL}_k(\mathbb{F}_{q^m})$, $\sigma \in \mathrm{GL}_{n+\hat{t}}(\mathbb{F}_q)$, $G \in \mathbb{F}_{q^m}^{k\times n}$, and $X \in \mathbb{F}_{q^m}^{k\times \hat{t}}$ be of rank a, and consider Loidreau's GGPT variant with public key*

$$\hat{G}_{\mathrm{pub}} = S[X \mid G]\sigma.$$

If Assumptions 5.4.1 and 5.4.2 are true, then we can break the Loidreau GGPT variant in polynomial time with high probability.

Proof. Let ℓ be as in (5.7) and $T\hat{G}_{\mathrm{pub}}$ as in (5.6). Consider the matrix

$$G''_{\mathrm{ext}} := \begin{pmatrix} \hat{G}_{\mathrm{pub}} \\ \hat{G}_{\mathrm{pub}}^{([k-a])} \\ \vdots \\ \hat{G}_{\mathrm{pub}}^{([(k-a)(\ell-1)])} \end{pmatrix} = \tilde{S} \underbrace{\begin{pmatrix} (X^* \mid G^*) \\ (X^* \mid G^*)^{([k-a])} \\ \vdots \\ (X^* \mid G^*)^{([(k-a)(\ell-1)])} \\ \hline (0 \mid G^{**}) \\ (0 \mid G^{**})^{([k-a])} \\ \vdots \\ (0 \mid G^{**})^{([(k-a)(\ell-1)])} \end{pmatrix}}_{\bar{G}} \sigma.$$

Since $\langle G^{**} \rangle_{\mathbb{F}_{q^m}}$ is a subcode of $\langle G \rangle_{\mathbb{F}_{q^m}}$ of codimension a, by Assumption 5.4.1, we have with high probability,

$$\sum_{i=0}^{\ell-1} \langle G^{**} \rangle^{([(k-a)i])} = \mathbb{F}_{q^m}^n.$$

Then, the bottom submatrix of $\bar{G}$ has the same row span as $[0 \mid I_n]$, and hence, by using elementary operations, we can eliminate the second component of every row in the top submatrix of $\bar{G}$. Then, the space generated by the rows of G''_{ext} is the same as that generated by

$$G'''_{\mathrm{ext}} = \left(\begin{array}{c|c} X^* & 0 \\ (X^*)^{([k-a])} & 0 \\ \vdots & \vdots \\ (X^*)^{([(k-a)(\ell-1)])} & 0 \\ 0 & I_n \end{array} \right) \sigma.$$

By Assumption 5.4.2, with high probability we have that

$$\sum_{i=0}^{\ell-1} \langle X^* \rangle_{\mathbb{F}_{q^m}}^{([(k-a)i])}$$

contains no elements of rank one, and therefore all elements of rank one in $\langle G''_{\mathrm{ext}} \rangle_{\mathbb{F}_{q^m}} = \langle G'''_{\mathrm{ext}} \rangle_{\mathbb{F}_{q^m}}$ must belong to $\langle [0 \mid I_n] \rangle_{\mathbb{F}_{q^m}} \sigma$. With the help of Lemma 5.2.11 we can recover a matrix $U \in \mathbb{F}_q^{n \times (\hat{t}+n)}$ which is a basis for $\langle [0 \mid I_n] \rangle_{\mathbb{F}_{q^m}} \sigma$. Then, any parity check matrix $H_U \in \mathbb{F}_q^{\hat{t} \times (n+\hat{t})}$ for $\langle [0 \mid I_n] \rangle \sigma$ must have the form

$$H_U^T = \sigma^{-1} \begin{bmatrix} A \\ 0 \end{bmatrix} \in \mathbb{F}_q^{(n+\hat{t}) \times \hat{t}},$$

where $A \in \mathrm{GL}_{\hat{t}}(\mathbb{F}_q)$. It follows that if we compute

$$\hat{G}_{\mathrm{pub}} H_U^T = SXA \in \mathbb{F}_{q^m}^{k \times \hat{t}},$$

then there exists a unique matrix $V = [A^{-1} \mid 0]\sigma \in \mathbb{F}_q^{\hat{t}\times(n+\hat{t})}$ such that

$$\hat{G}_{\text{pub}} H_U^T V = SXAV = S[X \mid 0]\sigma.$$

We can find the matrix V by observing that

$$(\hat{G}_{\text{pub}} - \hat{G}_{\text{pub}} H_U^T V) H_U^T = S[0 \mid G]\sigma H_U^T = 0. \tag{5.9}$$

This gives a linear system of equations with $\hat{t}(n+\hat{t})$ variables and $k\hat{t}$ equations over $\mathbb{F}_{q^m}$. Since the variables can take values in $\mathbb{F}_q$, we can expand each equation into m equations over $\mathbb{F}_q$, obtaining a system of $km\hat{t}$ equations and $\hat{t}(n+\hat{t})$ variables over $\mathbb{F}_q$. Hence, we can solve if $km \geq n + \hat{t}$ which is always satisfied for GGPT cryptosystems.

Now, let $H_V \in \mathbb{F}_{q^m}^{n\times(n+\hat{t})}$ be any dual matrix for V. Then, H_V has the form

$$H_V^T = \sigma^{-1} \begin{bmatrix} 0 \\ B \end{bmatrix},$$

for some $B \in \mathrm{GL}_n(\mathbb{F}_q)$. Therefore,

$$\hat{G}_{\text{pub}} H_V^T = SGB \in \mathbb{F}_{q^m}^{k\times n}$$

is a Gabidulin code of minimum distance $n-k+1$, from which we can recover a decoding algorithm, as in Lemma 5.2.1. If we receive an encrypted message of the form

$$\boldsymbol{m}\hat{G}_{\text{pub}} + \boldsymbol{e},$$

we can apply H_V^T, obtaining

$$\boldsymbol{m}SGB + \boldsymbol{e}H_V^T.$$

Since

$$\mathrm{rk}(\boldsymbol{e}H_V^T) \leq \mathrm{wt}_{\mathrm{R}}(\boldsymbol{e}) \leq n - k + 1,$$

we can recover the encrypted message, $\boldsymbol{m}$, from the recovered decoding algorithm with respect to SGB. All the operations required for this attack can be performed in polynomial time. □

We will conclude this section with an example in which we perform our attack against the parameters proposed by Loidreau in [49] in order to resist Overbeck's attack. The proposed parameters are not secure against our attack.

Example 5.4.4. Consider an LGGPT variant with $q = 2$, $m = n = 24$, $k = 12$, $a = 3$, and $\hat{t} = 40$, i.e. the first set of parameters from Table 5.1. Assume we, as an attacker, know the public generator matrix $\hat{G}_{\text{pub}} \in \mathbb{F}_{2^{24}}^{12\times 64}$ and received an encrypted message $\boldsymbol{y}$. We compute $\ell = \lceil \frac{24}{12-3} \rceil = 3$ and proceed as follows:

1. We compute $\hat{G}_{\text{pub}}^{([9])}, \hat{G}_{\text{pub}}^{([18])}$ to obtain the extended matrix $G''_{\text{ext}} \in \mathbb{F}_{2^{24}}^{36\times 64}$. This requires at most $1536 = 2 \cdot 12 \cdot 64$ Frobenius powers in $\mathbb{F}_{2^{24}}$. Using a normal basis to represent $\mathbb{F}_{2^{24}}$ over $\mathbb{F}_2$, this can be done very efficiently.

m	n	k	a	$\hat{t}$	Assumption 5.4.1	Assumption 5.4.2
24	24	12	3	40	~ 1	~ 1
24	24	12	4	52	$\sim .998$	~ 1

Table 5.1: Experimental results for Assumptions 1 and 2: Probabilities of success in 1000 trials for $q = 2$.

2. We find the elements of rank one in $\langle G''_{\text{ext}} \rangle_{\mathbb{F}_{q^m}}$, as described in Lemma 5.2.11. To do so we need to row reduce G''_{ext} and then solve a linear system over $\mathbb{F}_2$ with 36 unknowns and $24 \cdot 64 = 1536$ equations. Then, if Assumptions 1 and 2 hold, we find some basis matrix $U \in \mathbb{F}_2^{24\times 64}$, such that $\langle U \rangle_{\mathbb{F}_{q^m}}$ contains all these elements of rank one.

3. Compute a dual matrix H_U for U.

4. We find a matrix $V \in \mathbb{F}_2^{40\times 64}$, solving Equation (5.9).

5. We compute a parity check matrix $H_V \in \mathbb{F}_2^{24\times 64}$ for V, and compute the product $\hat{G}_{\text{pub}} H_V^T$.

6. We recover a decoding algorithm for the code $\hat{G}_{\text{pub}} H_V^T$, as described in Lemma 5.2.11, and decode $\boldsymbol{y} H_V^T$ with this algorithm.

We observe that step 4 above is the most computationally intensive, and therefore we estimate the complexity of our attack based on these step. This is done by solving a $(40 \cdot 64) \times (24 \cdot 12 \cdot 40)$ system over $\mathbb{F}_2$ by Gaussian elimination on the resulting matrix. This requires on the order of 2^{39} operations over $\mathbb{F}_2$. Implementing the algorithm on a personal computer, we are able to break random instances of this cryptosystem with high probability for the proposed parameters. Futhermore the attack appears to be resilient to changes in the values of the parameters.

Table 5.1 contains some preliminary results regarding the validity of Assumptions 5.4.1 and 5.4.2. For the parameters in the second row of the table, we can similarly break the system, albeit with slightly higher complexity due the larger parameters. We note that these parameters were previously deemed insufficient using generic attacks [30]. Nevertheless, our attack is significantly faster and would even be resilient against an increase in the parameters.

5.4.2 SA Variant Cryptanalysis

Recall that the designed error matrix in the SA variant is constructed as $X = X_{\text{Moore}} + Z$, where this decomposition is a minimal column rank Moore decomposition, and X_{Moore} has rank a and Z has column rank $\hat{t} - a$. We can rewrite

$$\hat{G}_{\text{pub}} = \underbrace{S[X_{\text{Moore}} \mid G]\sigma}_{M} + \underbrace{S[Z \mid 0]\sigma}_{X'}. \tag{5.10}$$

We observe that $S^{-1}M$ is a Moore matrix generating a code of minmum rank distance at least $n-k+1$, and X' is a matrix of column rank $\hat{t}-a$.

Theorem 5.4.5. *Consider a GGPT cryptosystem as defined above. Suppose an adversary can find a matrix $U' \in \mathbb{F}_q^{(\hat{t}-a)\times(\hat{t}+n)}$ such that $\langle U'\rangle_{\mathbb{F}_{q^m}} = \mathrm{supp}_{\mathrm{Gr}}([Z \mid 0]\sigma) = \mathrm{supp}_{\mathrm{Gr}}(X')$. Then an encrypted message from a public key of the form (4.4) can be recovered in polynomial time.*

Proof. Let $U \in \mathbb{F}_q^{(\hat{t}-a)\times\hat{t}}$ be a Grassmann support matrix for Z, that is $\langle U\rangle_{\mathbb{F}_{q^m}} = \mathrm{supp}_{\mathrm{Gr}}(Z)$. Then, $U' = [U \mid 0]\sigma$ is a Grassmann support matrix for X'. Let $H_U \in \mathbb{F}_q^{a\times\hat{t}}$ be a parity check matrix for U. If $H_{U'} \in \mathbb{F}_q^{(a+n)\times(\hat{t}+n)}$ is any parity check matrix for U', then $(H_{U'})^T$ must be of the form

$$(H_{U'})^T = \sigma^{-1}\left[\begin{array}{c|c} H_U^T & 0_{\hat{t}\times n} \\ \hline 0_{n\times a} & I_n \end{array}\right] A,$$

for some $A \in \mathrm{GL}_{n+a}(\mathbb{F}_q)$.

We compute,

$$\begin{aligned}\hat{G}_{\mathrm{pub}}(H_{U'})^T &= S[X_{\mathrm{Moore}} \mid G]\sigma(H_{U'})^T + S[Z \mid 0]\sigma(H_{U'})^T \\ &= S[X_{\mathrm{Moore}} \mid G]\left[\begin{array}{c|c} H_U^T & 0_{\hat{t}\times n} \\ \hline 0_{n\times a} & I_n \end{array}\right] A \\ &= S[X_{\mathrm{Moore}}H_U^T \mid G]A.\end{aligned}$$

Since $X_{\mathrm{Moore}}H_U^T$ is again a Moore matrix, we can find n independent columns of $\hat{G}_{\mathrm{pub}}(H_{U'})^T$ which will form a Gabidulin code of minimum distance $n-k+1$, which has error correction capability t'. Denote the columns by $\boldsymbol{i} = (i_1, \ldots, i_n)$ and the corresponding submatrix, $G_{\boldsymbol{i}}$. From Lemma 5.2.1, we can recover a decoding algorithm for $\langle G_{\boldsymbol{i}}\rangle_{\mathbb{F}_{q^m}}$ with respect to $G_{\boldsymbol{i}}$. We note that if $\boldsymbol{e}$ is an error of rank at most t', and we denote by $\boldsymbol{e'}$ the subvector of $\boldsymbol{e}(H_{U'})^T$ corresponding to columns $\boldsymbol{i}$, then

$$\mathrm{rk}(\boldsymbol{e'}) \leq \mathrm{rk}(\boldsymbol{e}(H_{U'})^T) \leq \mathrm{rk}(\boldsymbol{e}) \leq t'.$$

If $\boldsymbol{m}$ is an encrypted message of the form $\boldsymbol{m} = \boldsymbol{m}\hat{G}_{\mathrm{pub}} + \boldsymbol{e}$, then applying $H_{U'}^T$ we obtain

$$\boldsymbol{m}(H_{U'})^T = \boldsymbol{m}\hat{G}_{\mathrm{pub}}(H_{U'})^T + \boldsymbol{e}(H_{U'})^T.$$

Restricting to the coordinates $\boldsymbol{i}$, we obtain

$$\boldsymbol{m}G_{\boldsymbol{i}} + \boldsymbol{e'},$$

from which we can decode to recover $\boldsymbol{m}$. □

Corollary 5.4.6. *We can recover an encrypted message from the SA variant of the GGPT cryptosystem in polynomial time if*

$$\hat{t} - a < \frac{n-k-1}{2}.$$

Proof. Recall that

$$\hat{G}_{\text{pub}} = \underbrace{S[X_{\text{Moore}} \mid G]\sigma}_{M} + \underbrace{S[Z \mid 0]\sigma}_{X'},$$

and note that $S^{-1}M + S^{-1}X'$ is a minimal column rank Moore decomposition of $S^{-1}\hat{G}_{\text{pub}}$. $S^{-1}M$ is a Moore matrix generating a code of minimum rank distance at least $n-k+1$. Since

$$n - k + 1 > 2(\hat{t} - a) + 2,$$

from Corollary 5.2.10, all elements of rank one in

$$\sum_{i=0}^{\hat{t}-a} \langle S^{-1}M + S^{-1}X' \rangle_{\mathbb{F}_{q^m}}^{([i])}$$

span the space $\text{supp}_{\text{Gr}}(X') = \text{supp}_{\text{Gr}}([Z \mid 0]\sigma)$. We can then find these elements of rank one using Corollary 5.2.10, to obtain a matrix $U' \in \mathbb{F}_q^{(\hat{t}-a)\times(\hat{t}+n)}$ such that $\langle U' \rangle_{\mathbb{F}_{q^m}} = \text{supp}_{\text{Gr}}(X')$. Applying Theorem 5.4.5, we can break the SA variant of the GGPT cryptosystem. □

The following example illustrates a small example of how Overbeck's attack fails, but our attack recovers the encrypted message.

Example 5.4.7. Let $q = 2$, $n = 8$, $k = 3$, $\hat{t} = 3$, $a = 1$ and $g_1, \ldots, g_8 \in \mathbb{F}_{2^8}$ linearly independent over $\mathbb{F}_2$. Consider the generator matrix of a Gabidulin code

$$G = \begin{pmatrix} g_1 & g_2 & \cdots & g_8 \\ g_1^{([1])} & g_2^{([1])} & \cdots & g_8^{([1])} \\ g_1^{([2])} & g_2^{([2])} & \cdots & g_8^{([2])} \end{pmatrix},$$

and, for some $x \in \mathbb{F}_{2^8} \backslash \mathbb{F}_2$, the matrices

$$X_{\text{Moore}} = \begin{pmatrix} x & 0 & 0 \\ x^{([1])} & 0 & 0 \\ x^{([2])} & 0 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Let

$$X = X_{\text{Moore}} + Z,$$

so that the public key generator matrix is given by

$$\hat{G}_{\text{pub}} = [X \mid G] = \left(\begin{array}{ccc|ccc} x & 1 & 1 & g_1 & g_2 & \cdots & g_8 \\ x^{([1])} + 1 & 0 & 1 & g_1^{([1])} & g_2^{([1])} & \cdots & g_8^{([1])} \\ x^{([2])} + 1 & 1 & 0 & g_1^{([2])} & g_2^{([2])} & \cdots & g_8^{([2])} \end{array}\right).$$

For simplicity we let $S = I_3$ and $\sigma = I_{11}$. Compute the extended matrix,

$$G_{\text{ext}} = \begin{pmatrix} \hat{G}_{\text{pub}} \\ \hat{G}_{\text{pub}}^{([1])} \end{pmatrix}$$

which can be put in the form

$$G'_{\text{ext}} = \left(\begin{array}{ccc|cccc} x & 0 & 0 & g_1 & g_2 & \dots & g_8 \\ x^{([1])} & 0 & 0 & g_1^{([1])} & g_2^{([1])} & \dots & g_8^{([1])} \\ x^{([2])} & 0 & 0 & g_1^{([2])} & g_2^{([2])} & \dots & g_8^{([2])} \\ x^{([3])} & 0 & 0 & g_1^{([3])} & g_2^{([3])} & \dots & g_8^{([3])} \\ 1 & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 1 & 0 & 0 & \dots & 0 \end{array}\right) = \left(\begin{array}{c|c} X^* & G^* \\ X^{**} & 0 \end{array}\right)$$

by a suitable row transformation. Here Overbeck's attack fails, because X^{**} does not have full rank. On the other hand, our attack succeeds, since we can directly recover the elements of rank one as $\langle [X^{**} \mid 0] \rangle_{\mathbb{F}_{q^m}} = \langle [Z \mid 0] \rangle_{\mathbb{F}_{q^m}}$. Thus we can use Theorem 5.4.5 and recover any encrypted message.

5.5 Column Scrambler Variant and Generalization of Gaborit's Attack

Recall the CS variant from Subsection 4.2.4. Let G be a generator matrix for a t-error correcting Gabidulin code. Let $0 < t_1 < t$ be a design parameter and let $\mathcal{P}_{n,t,t_1}(\mathbb{F}_{q^m}) \subset \mathrm{GL}_n(\mathbb{F}_{q^m})$ be given by

$$\begin{aligned} \mathcal{P}_{n,t,t_1} = \{[Q_1 \mid Q_2]\sigma \mid & Q_1 \in \mathbb{F}_{q^m}^{n\times(t-t_1)}, Q_2 \in \mathbb{F}_q^{n\times(n-t+t_1)} \\ & \text{s.t. } \mathrm{rk}([Q_1 \mid Q_2]) = n, \sigma \in \mathrm{GL}_n(\mathbb{F}_q)\}. \end{aligned}$$

For a random $P \in \mathcal{P}_{n,t,t_1}(\mathbb{F}_{q^m})$, the public generator matrix is of the form $\kappa_{\text{pub}} = SGP^{-1}$.

The structure of P^{-1} is not immediately clear from the definition of $\mathcal{P}_{n,t,t_1}(\mathbb{F}_{q^m})$. The assumption behind the security of such a system is that the matrices P^{-1} appear random in the set $\mathrm{GL}_n(\mathbb{F}_{q^m})$. We show that by considering carefully the elements of rank one in the extended code, the matrices $P \in \mathcal{P}$ are revealed to have quite a lot of structure. In particular, they do not alter the structure of G sufficiently enough so that we can recover some information from the elements of rank one in the extended code. While we do not prove any results about the structure of P^{-1}, we generate data to support some assumptions which, if true, allow us to easily break the CS system. The attack on this particular variant is therefore unlike the cryptanalysis of the LGGPT and SA variants, where we can exhibit a proof of a break. Rather than find a structural flaw in P^{-1} that renders it amenable to an attack, we instead give statistical evidence that there exists a exploitable structure of the public key in the generic case. In fact, in thousands of randomly generated instances of the column scrambler variant, there were none in which our attack did not succeed.

Recall in Proposition 5.1.2 that it was shown how generic rank metric codes with certain parameters sets were found to have polynomial time decoding algorithms. In order to emphasize the difference in our approach, we recall that the proof required the use of linearized polynomials and argued based on the ability to reconstruct the vector space support of the error vector. We will take an alternative approach which instead

considers the Grassmann support of the error vector, which works for more general parameters than Proposition 5.1.2.

Theorem 5.5.1. *Let $\mathcal{C}$ be a code with parity check matrix $H \in \mathbb{F}_{q^m}^{(n-k)\times n}$ capable of correcting any error pattern, $\boldsymbol{e}$, of weight r. Let $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}$ where $\boldsymbol{x} \in \mathcal{C}$. Suppose that for some s, with $(s, m) = 1$, we have*

$$\dim(\mathcal{C}^{([s])} \cap \mathcal{C}) = \ell.$$

If

$$(k+1)(r+1) - 1 \leq n + (r-1)\ell \tag{5.11}$$

and $\mathcal{C} \cap \mathcal{U} = \{0\}$, where $\mathcal{U}$ is the space spanned by the elements of rank one in

$$\mathcal{C}_{\text{ext}} = \sum_{i=0}^{r-1} \left(\mathcal{C} + \langle \boldsymbol{e} \rangle_{\mathbb{F}_{q^m}}\right)^{([si])},$$

then

$$\mathcal{R}_{q^m,n,k}(H, H\boldsymbol{y}^T, r) \sim O(r(k+1)mn^2).$$

Proof. Let $G \in \mathbb{F}_{q^m}^{k\times n}$ be a generator matrix for $\langle H \rangle_{\mathbb{F}_{q^m}}^{\perp}$ and $\boldsymbol{m}G \in \mathcal{C}$. Let $\boldsymbol{y} = \boldsymbol{m}G + \boldsymbol{e}$ be a corrupted codeword. Observe that

$$\text{supp}_{\text{Gr}}(\boldsymbol{e}) = \sum_{i=0}^{r-1} \langle \boldsymbol{e} \rangle_{\mathbb{F}_{q^m}}^{([si])} \subseteq \mathcal{U}.$$

Denote $u = \dim(\mathcal{U})$, and let $H_U \in \mathbb{F}_q^{(n-u)\times n}$ be a parity check matrix for $\mathcal{U}$. We have,

$$u \leq \dim(\mathcal{C}_{\text{ext}}) \leq (k+1)r - (r-1)\ell \leq n - k,$$

where the last inequality follows from (5.11). Moreover, we have

$$\boldsymbol{y}H_U^T = \boldsymbol{m}GH_U^T + \boldsymbol{e}H_U^T = \boldsymbol{m}GH_U^T. \tag{5.12}$$

Since $\mathcal{C} \cap \mathcal{U} = \{0\}$, we have that that GH_U^T is a full rank matrix. Therefore we can solve (5.12) in polynomial time, e.g. with the Gaussian algorithm, which implies the statement. In order to obtain the complexity, we assume that taking Frobenius powers is computationally inexpensive by supposing we represent vectors in $\mathbb{F}_{q^m}^n$ via a normal basis. Then, the most expensive step is to obtain H_U, which we can do by Lemma 5.2.11, where we need to find the elements of rank one in the span of a $r(k+1) \times n$ matrix which represents $\mathcal{C}_{\text{ext}}$. □

Remark 5.5.2. The condition (5.11) in Theorem 5.5.1 is not necessary, however, it allows for a comparison to the conditions in Proposition 5.1.2. We note that when $\ell = 0$, this bound reduces to that obtained by Gaborit et al. To generalize this idea, we note that this attack can be applied as long as the condition $\text{rk}(GH_U^T) = k$ holds; i.e. the only condition we actually require is that $\mathcal{C} \cap \mathcal{U} = \{0\}$. This raises a similar concern as in the proof of 5.1.2 in which it is necessary for the system arising in (5.1) to be a full rank system of equations. It is not clear how the condition that GH_U^T being full rank (equivalently $\mathcal{C} \cap \mathcal{U} = \{0\}$) relates to the condition of (5.1) being full rank.

In the proof of Theorem 5.5.1, the bound (5.11) arises from the worst case scenario, $u = \dim(\mathcal{C}_{\text{ext}})$. In general this bound may be quite inaccurate. We can actually use a weaker condition to guarantee that we can still break the cryptosystem. We therefore make the following definition.

Definition 5.5.3. Let $\mathcal{C} \in \mathbb{F}_{q^m}^N$ be a linear code of dimension k with rank error correction capability r. We say that $\mathcal{C}$ is *r-Frobenius weak* if, for some s relatively prime to m and for any $\boldsymbol{e} \in \mathbb{F}_{q^m}^N$ of rank at most r, the space $\mathcal{U}$, spanned by the elements of rank one in

$$\sum_{i=0}^{r-1} (\mathcal{C} + \langle \boldsymbol{e} \rangle_{\mathbb{F}_{q^m}})^{([si])} = \sum_{i=0}^{r-1} \mathcal{C}^{([si])} + \text{supp}_{\text{Gr}}(\boldsymbol{e}),$$

satisfies $\mathcal{C} \cap \mathcal{U} = \{0\}$.

Remark 5.5.4. If $\mathcal{C}$ is r-Frobenius weak, then we can solve the rank syndrome decoding problem for $\mathcal{C}$ in polynomial time. In general, we do not need such a certain condition; it is enough that for most choices of $\boldsymbol{e} \in \mathbb{F}_{q^m}^N$ of rank r, $\dim(\mathcal{C} \cap \mathcal{U}) = \{0\}$. Then, such a code would be a bad candidate for a rank-based cryptosystem.

We now return to the CS variant. Table 5.2 summarizes some values for the CS variant, including the rank of $G_{\text{pub}} H_U^T = (SGP^{-1}) H_U^T$, where H_U is as in the proof of Theorem 5.5.1. These values were generated from random instances of the CS variant. One can then deduce – without explicitly describing their structure – that the matrices SGP^{-1} are rather non-generic. It appears that they are t_1-Frobenius weak, probably because they have quite high intersection with their coordinate-wise Frobenius image. Based on experimental data, we make the following assumption.

Assumption 5.5.5. *A matrix of the form SGP^{-1} as in 4.2.4 is t_1-Frobenius weak.*

Corollary 5.5.6. *If Assumption 5.5.5 holds, then we can break the CS variant in polynomial time.*

For all instances of the values we tested, Assumption 5.5.5 appears to hold. In particular, in thousands of randomly generated encrypted messages according to the CS variant, there were no instances in which the attack did not successfully recover the message. The necessary code is given in Appendix A.3. Therefore, we consider the CS variant unusable. As explained before, for certain parameters, Gaborit et al. were able to cryptanalyze the CS variant using a generic rank syndrome decoding algorithm [30]. However, a small increase in the parameters renders the generic algorithm infeasible. In contrast, our algorithm remains effective even for larger parameters, as evidenced by Table 5.2.

Remark 5.5.7. In this chapter, we attacked and manage to cryptanalyze nearly all existing variants of rank metric based cryptosystems where the underlying code is Gabidulin. Our attack generalizes the attack of Overbeck, and is based on computing the elements of rank one in an extended code. Using this same idea, we are able to establish a generalization of a result by Gaborit et al. which indicates that codes that overlap with

m	n	k	t_1	u	$\mathrm{rk}(G_{\mathrm{pub}} H_U^T)$
28	28	14	3	7	14
28	28	14	4	7	14
42	42	21	4	11	21
42	42	21	5	11	21
42	42	21	6	11	21
48	48	24	6	12	24
48	48	24	7	12	24
48	48	24	8	12	24
48	48	24	9	12	24

Table 5.2: Experimental results for the CS variant

respect to the coordinate-wise Frobenius map may be weak candidates for use in rank based cryptography.

We note then, that codes also based on the twisted Gabidulin codes of Sheekey would also be poor choices for use in rank metric based cryptosystems, since their intersection is almost as large as that of Gabidulin codes (they intersect their coordinate-wise Frobenius power in dimension just one less). On the other hand, the LRPC codes introduced in Subsection 3.2.3 do not in general intersect their Frobenius powers. Therefore, they may resist such attacks as described here. However, other structural flaws may be found which may allow one to exploit the elements of rank one. In general, more attention is needed to determine if these codes resist all known attacks, including attacks based on creating elements of rank one as outlined in this thesis. To the best of this author's knowledge, after cryptanalysis of the GGPT and CS variants, the LRPC codes are the only remaining feasible proposal for rank metric based cryptography.

Chapter 6

Subspace Fuzzy Vault

Fuzzy vault is the term used by Juels and Sudan in [42] to describe a cryptographic primitive in which a key, κ, is hidden by a set of features, A, in such a way that any witness, B which is close enough to A under the set difference metric can decommit κ. Fuzzy vault is related to the fuzzy commitment scheme of Juels and Wattenberg [41], as well as the notion of fuzzy extractors [22]. The motivation for fuzzy vault is related to the growing interest in using fuzzy authentication systems, i.e. systems that do not require an exact key match, but rather a partial one. An obvious application of such a system would be a biometric authentication scheme, since one cannot expect a biometric submission to be replicated exactly every time. Other applications include personal entropy systems and privacy-protected matching.

In early biometric systems, comparison of the biometric was performed against an image stored locally in unencrypted form so an image processing algorithm could compare directly the images. For security purposes–as is the case, for instance, with passwords– sensitive data should be stored in some indecipherable form. In the case of passwords, a significantly complicated hash function is used, however, hash values are dramatically different even when the values of the inputs are very close. Therefore, hashing of fingerprints and comparison of hash values would make it difficult to compare an authentic user with a stored template. Without performing some sort of hashing, sensitive data can be easily read if an attacker gains unintended access. In the case of biometric data, an adversary with access to the biometric has the same access as the authentic user; it is unsafe to use again on any system. Therefore, it is of utmost importance to secure biometric data.

The main results of this chapter have recently appeared in [53] in collaboration with the coauthors. In Section 6.1, we present the fuzzy vault scheme proposed by Juels and Sudan, which works for codes that can be represented as the evaluation of polynomials. In Section 6.2, we propose an alternative construction of the fuzzy vault designed to work for constant-dimension subspace codes. In doing so, there are some extra concerns that much be taken. Although estimates for security are somewhat similar, using subspace codes allows for a different trade-off in parameter relationships for a fuzzy vault system. The security and other concerns are investigated in Section 6.3.

6.1 Preliminaries

As was mentioned briefly in Section 2.3, we want to create a locking system that releases the key if the difference between the template and a witness is small. In order to measure this difference, we will use the following metric, which we note is permutation invariant.

Definition 6.1.1. Let A, B be sets. The *set difference metric*, $d_{\triangle}$, is given by

$$d_{\triangle}(A, B) = |(A \setminus B) \cup (B \setminus A)|.$$

We briefly state some of the necessary background and then define the fuzzy vault as it is given in [42], which we will distinguish from our construction by referring to it as the *polynomial fuzzy vault* (PFV) scheme.

Definition 6.1.2. Let $\mathcal{C} \subset \mathbb{F}_{q^m}^n$ be a rank metric code. The subspace code defined by

$$L(\mathcal{C}) = \{\langle [I_m \mid [\boldsymbol{x}]] \rangle_{\mathbb{F}_q} \mid \boldsymbol{x} \in \mathcal{C}\} \subset \mathrm{Gr}(m, \mathbb{F}_q^{n+m}),$$

is called the *lifting* of the rank metric code, $\mathcal{C}$.

For more information on construction of subspace codes by lifting of rank metric codes, the reader is directed to [68, 23, 74]. The following lemma can be found in any of the aforementioned papers.

Lemma 6.1.3. *Let $\mathcal{C} \subset \mathbb{F}_{q^m}^n$ be a rank metric code of minimum distance d and dimension k. Then,*

$$d_{\mathrm{S}}^{\min}(L(\mathcal{C})) = 2d.$$

Proof. Let $[\boldsymbol{x}], [\boldsymbol{y}] \in \mathbb{F}_q^{m \times n}$ for codewords $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$. Note that $\dim(\langle [I_m \mid [\boldsymbol{x}]] \rangle_{\mathbb{F}_q} = m$ for every $\boldsymbol{x} \in \mathcal{C}$. We have

$$\begin{aligned}
\dim(\langle [I_m \mid [\boldsymbol{x}]] \rangle_{\mathbb{F}_q} + \langle [I_m \mid [\boldsymbol{y}]] \rangle_{\mathbb{F}_q}) &= \mathrm{rk} \left(\begin{array}{c|c} I_m & [\boldsymbol{x}] \\ I_m & [\boldsymbol{y}] \end{array} \right) \\
&= \mathrm{rk} \left(\begin{array}{c|c} I_m & [\boldsymbol{x}] \\ 0 & [\boldsymbol{y} - \boldsymbol{x}] \end{array} \right) \\
&\geq m + d_{\mathrm{S}}^{\min}(\mathcal{C}).
\end{aligned}$$

Also,

$$\dim(\langle [I_m \mid [\boldsymbol{x}]] \rangle_{\mathbb{F}_q} \cap \langle [I_m \mid [\boldsymbol{y}]] \rangle_{\mathbb{F}_q}) \leq m - d_{\mathrm{S}}^{\min}(\mathcal{C}).$$

Therefore,

$$d_{\mathrm{S}}(\langle [I_m \mid [\boldsymbol{x}]] \rangle_{\mathbb{F}_q}, \langle [I_m \mid [\boldsymbol{y}]] \rangle_{\mathbb{F}_q}) \geq 2d_{\mathrm{S}}^{\min}(\mathcal{C})$$

with equality if $\boldsymbol{x}, \boldsymbol{y}$ attain the minimum distance of $\mathcal{C}$. □

Recall from Definition 2.1.3, that a subspace code $\mathcal{S} \subset \mathrm{Gr}_q(k, \mathbb{F}_q^n)$ forms a spread if the codewords of $\mathcal{S}$ intersect trivially and the union of the codewords is all of $\mathbb{F}_q^n$. Such a subspace code will be called a *spread code*. Let $\boldsymbol{\alpha} \in \mathbb{F}_{q^k}^k$ be a vector of independent elements over $\mathbb{F}_q$ and let $\mathcal{C}$ be the Gabidulin code $\mathrm{Gab}_{k,1}(\boldsymbol{\alpha})$. Each element $\boldsymbol{x} \in \mathbb{F}_{q^k}^k$ corresponds to a matrix $[\boldsymbol{x}] \in \mathbb{F}_q^{k \times k}$. Since the minimum distance of $\mathcal{C}$ is k, then $L(\mathcal{C})$ has minimum distance $2k$ (hence the codewords of $L(\mathcal{C})$ intersect trivially). Let

$$\mathcal{S} = L(\mathcal{C}) \cup \{\langle [0 \mid I_k] \rangle_{\mathbb{F}_q}\}. \tag{6.1}$$

To see that $\mathcal{S}$ is a spread code, we first observe that

$$|S| = q^k + 1 = \frac{q^{2k} - 1}{q^k - 1},$$

so S has the correct cardinality. Furthermore, we can observe that

$$\mathrm{rk} \left(\begin{array}{c|c} I_k & [\boldsymbol{x}] \\ 0 & I_k \end{array} \right) = 2k$$

for every $\boldsymbol{x} \in \mathcal{C}$. Together with Lemma 6.1.3 we see that $\mathcal{S}$ is indeed a spread code.

Inductively, we can create spread codes for any n, k, q, as long as $n = k\ell$ is a multiple of k, in the following way. For $i = 1, \ldots, \ell$, define

$$S_{k,\ell,i} = \{\langle [\underbrace{0_{k\times k}}_{\ell\text{th block}} \mid \ldots \mid 0_{k\times k} \mid \underbrace{I_k}_{i\text{th block}} \mid [\boldsymbol{x}_{i-1}] \mid \ldots \mid [\boldsymbol{x}_1]] \rangle_{\mathbb{F}_q} \mid \boldsymbol{x}_j \in \mathcal{C}\}.$$

Then, the subspace code

$$\mathcal{S}_{k,\ell} = \bigcup_{i=1}^{\ell} \mathcal{S}_{k,\ell,i}$$

is a $(k\ell, k)_q$-spread code. Note that we have already constructed $\mathcal{S}$ from (6.1) as

$$\mathcal{S}_{k,2} = \mathcal{S}_{k,2,1} \cup \mathcal{S}_{k,2,2}.$$

This construction is equivalent to the construction of [52]. Generalizations of spreads for the case when $k \nmid n$ have been considered, for instance in [36].

Spread codes are not the only constant dimension subspace codes. Other examples include codes from Ferrer diagrams [24] and a Reed-Solomon-like construction [44].

We now describe the PFV scheme. Let $\boldsymbol{\kappa} = (\kappa_0, \ldots, \kappa_{\ell-1}) \in \mathbb{F}_q^\ell$ be a secret key, and let

$$\kappa(x) = \sum_{i=0}^{\ell-1} k_i x^i, \tag{6.2}$$

be the corresponding key polynomial. Let $A \subset \mathbb{F}_q^*$ be a template with $|A| = t > \ell$. Choose $r > t$ and select a set $B \subset \mathbb{F}_q^* \setminus A$ such that $|B| = r - t$. Let $\lambda \colon \mathbb{F}_q \to \mathbb{F}_q$ be a random map such that $\lambda(x) \neq \kappa(x)$ for all $x \in B$. Construct the sets

$$\mathcal{P}_{\text{auth}} = \{(x, \kappa(x)) \mid x \in A\},$$

$$\mathcal{P}_{\text{chaff}} = \{(x, \lambda(x)) \mid x \in B\},$$

and set

$$\mathcal{V} = \mathcal{P}_{\text{auth}} \cup \mathcal{P}_{\text{chaff}}.$$

We will call $\mathcal{P}_{\text{auth}}$ the *authentic* points and the set $\mathcal{P}_{\text{chaff}}$ the *chaff* points. V will be called the *vault*.

Let $\boldsymbol{g} = (g_1, \ldots, g_t)$ be some ordering of the points in A, and construct the Reed-Solomon code, $\mathcal{C} = \text{RS}_{t,\ell}(\boldsymbol{g})$ as in Definition 2.0.10. Then, the evaluation of the key polynomial on the points $\boldsymbol{g}$ is a codeword of $\mathcal{C}$. In order to access the key, a witness submits a set of features, $W \subset \mathbb{F}_q$. Let $Z \subset \mathcal{V}$ be the set of vault points (x, y) with $x \in W$. As the error correction capability of $\mathcal{C}$ is $\lfloor (t-\ell)/2 \rfloor$, the witness needs

$$|Z \cap \mathcal{P}_{\text{auth}}| \geq t - \left\lfloor \frac{t-\ell}{2} \right\rfloor = d_{\text{H}}^{\min}(\mathcal{C}) - 1$$

in order to recover $\kappa(x)$ from a minimum distance decoding algorithm for $\mathcal{C}$.

The following attack was used by Mihăilescu in [55] to estimate the security of such a system. Since any witness set will result in a possibly spurious key, in order to attack the system, we need some guarantee that we have found the actual key. The idea behind the attack is to randomly build key polynomials by interpolating more than ℓ points. The probability that $\delta > \ell$ points interpolate to a polynomial of degree smaller than δ quickly tends to 0 under randomness assumptions. Therefore, if an attacker constructs a polynomial of smaller degree than the number of points used, it is fairly plausible that the attacker has stumbled upon the actual key.

We will assume that an attacker has access to the vault, but cannot distingiush the authentic and chaff sets. An adversary chooses a value $\ell < \delta \leq t$ and randomly generates δ elements of $\mathcal{V}$. Say $W' = \{w_1, \ldots, w_\delta\}$ is such a set. The following refinement of this idea uses the following lemma, which is due to [15, 55].

Lemma 6.1.4. *Let $\ell \leq \delta \leq t$ be such that among all polynomials, g, of degree $\ell - 1$ arising from interpolation of some vault points, $\kappa(x)$ is the only one which interpolates at least δ points with probability close to 1. Then, κ can be recovered in less than $8\ell \log^2 \ell \cdot (r/t)^\ell$ operations.*

Using this result, it was shown that certain reasonable parameters for the PFV scheme cause the system to be susceptible to a brute force attack. Choi et al. in [14] sped up the attack by using a fast polynomial reconstruction algorithm. For further information, including concerns about implementation and feature alignment the reader is directed to [58, 78].

6.2 SFV Scheme

This particular variant of the fuzzy vault will be called the *subspace fuzzy vault* (SFV) in order to distinguish from Juels and Sudan's version which makes use of polynomial reconstruction. Unlike the PFV scheme in which the key is given by the coefficients of a

polynomial, the key $\hat{\kappa}$ in this scheme will be a subspace encoded via a matrix, κ, whose row space is $\hat{\kappa}$. We remark that our construction is analogous to the construction of the PFV in that the key is a codeword and the features encode (possibly redundant) information about the codeword.

Let $\ell \leq k \leq n$, $\mathcal{C} \subset \mathrm{Gr}(\ell, \mathbb{F}_q^n)$ be a constant dimension subspace code, and $\hat{\kappa} \in \mathcal{C}$ a secret subspace. Choose some $\kappa \in \mathbb{F}_q^{k \times n}$ such that $\mathrm{rowsp}(\kappa) = \hat{\kappa}$. We will hide the key by a set of features, $A \subset \mathbb{F}_q^k$ with $|A| = t \geq \ell$ and a set $B \subset \mathbb{F}_q^k \setminus A$. Let $\lambda \colon \mathbb{F}_q^k \to \mathbb{F}_q^n$ be a random map such that $\lambda(\boldsymbol{x}) \notin \hat{\kappa}$ for all $\boldsymbol{x} \in B$. Define the sets,

$$\mathcal{P}_{\text{auth}} = \{(\boldsymbol{x}, \boldsymbol{x}\kappa) \mid \boldsymbol{x} \in A\},$$
$$\mathcal{P}_{\text{chaff}} = \{(\boldsymbol{x}, \lambda(\boldsymbol{x})) \mid \boldsymbol{x} \in B\},$$

and

$$\mathcal{V} = \mathcal{P}_{\text{auth}} \cup \mathcal{P}_{\text{chaff}}.$$

For a set $S \subset \mathbb{F}_q^k$, we will denote by $\langle S \rangle_\kappa$ the subspace spanned by the elements $\{\boldsymbol{s}\kappa \mid \boldsymbol{s} \in S\}$ and $\langle S \rangle_\lambda$ the set spanned by the elements $\{\lambda(\boldsymbol{s}) \mid \boldsymbol{s} \in S\}$. In order for a witness to decommit $\hat{\kappa}$, a set $W \subset \mathbb{F}_q^k$ is submitted and the second coordinates of the elements in the vault whose first coordinates correspond to W are used to generate a subspace, $\hat{W}$. This subspace is then decoded to yield a codeword of $\mathcal{C}$.

We will require the following assumptions. Firstly, all points associated with the vault are randomly chosen. Secondly, an authentic witness always submits a set of features, W, such that $|W \setminus A|$ and $|A \setminus W|$ are small relative to ℓ. With the randomness assumption from earlier, this would imply that with high probability, $\langle W \setminus A \rangle_\lambda = |W \setminus A|$ and $\langle A \setminus W \rangle_\kappa = |A \setminus W|$. Furthermore, $\langle A \setminus W \rangle_\kappa \cap \langle W \setminus A \rangle_\lambda = \{0\}$.

Theorem 6.2.1. *Let $A, W \subset \mathbb{F}_q^k$ be the authentic and witness sets respectively. Under the assumptions above, an authentic witness can recover the key if*

$$d_\triangle(A, W) \leq \frac{1}{2}(d_{\mathrm{S}}^{\min}(\mathcal{C}) - 1).$$

Proof. We can express $\hat{W} = (\hat{W} \cap \hat{\kappa}) \oplus E$, for some subspace $E \subset \mathbb{F}_q^n$. Thus, we can uniquely recover $\hat{\kappa}$ from $\hat{W}$ if

$$d_{\mathrm{S}}(\hat{W}, \hat{\kappa}) \leq \frac{1}{2}(d_{\mathrm{S}}^{\min}(\mathcal{C}) - 1).$$

We obtain

$$\begin{aligned} d_\triangle(A, W) &= |W \setminus A| + |A \setminus W| \\ &= \dim(\langle W \setminus A \rangle_\lambda) + \dim(\langle A \setminus W \rangle_\kappa) \\ &= \dim(\langle W \setminus A \rangle_\lambda) + k - \dim(\langle A \cap W \rangle_k) \\ &= \dim(E) + k - \dim(\hat{k} \cap \hat{W}) \\ &\leq \frac{1}{2}(d_{\mathrm{S}}^{\min}(\mathcal{C}) - 1). \end{aligned}$$

The second equality follows the assumptions regarding authentic witnesses, and the fourth equality follows since $E = \langle W \setminus A \rangle_\lambda$ and $\langle A \setminus W \rangle_\kappa \cap \langle W \setminus A \rangle_\lambda = \{0\}$. □

The assumptions made for Theorem 6.2.1 to hold are generic assumptions that are expected for an authentic witness. If W is not an authentic witness, we must consider the possibility of a false positive.

We first note that it is not an advantage for an attacker to submit a large witness set, W. Even if $\hat{W}$ contains the key space, i.e. $\hat{W} \supset \hat{\kappa}$, we would have

$$\begin{aligned} d_{\mathrm{S}}(\hat{W}, \hat{\kappa}) &= \dim(\hat{W} + \hat{\kappa}) - \dim(\hat{W} \cap \hat{\kappa}) \\ &= \dim(\hat{W}) - \dim(\hat{\kappa}). \end{aligned}$$

Therefore, if

$$\dim(\hat{W}) > \ell + \frac{1}{2}(d_{\mathrm{S}}^{\min}(\mathcal{C}) - 1), \tag{6.3}$$

we cannot recover the key. This gives a bound for the size of any useful witness set since any witness of larger size will necessarily also have larger distance.

A witness, W, is a *false positive* if

$$d_{\mathrm{S}}(\hat{W}, \hat{\kappa}) \leq \frac{1}{2}(d_{\mathrm{S}}^{\min}(\mathcal{C}) - 1)$$

and

$$d_{\triangle}(W, A) > \frac{1}{2}(d_{\mathrm{S}}^{\min}(\mathcal{C}) - 1).$$

We always have the possibility of false positives because of the dependence relations inherent in working with subspaces. A false positive can occur if

1. $\dim(\hat{W}) < |W|$,
2. $\dim(\hat{\kappa}) < |A|$, or
3. $\dim(\hat{W} \cap \hat{\kappa}) > |W \cap A|$.

Under randomness assumptions, for large n and small witness size (recall, a large witness is not of value for an attacker), the probability of (1) is negligable. That is, we expect randomly chosen points of the vault to be independent if the number of points chosen is less than the dimension of the space (for instance one can invoke Lemma 6.3.1 in the following section). (2) is a design concern which can be mitigated by taking $\dim(\hat{\kappa})$ close to $|A|$. Again under randomness assumptions, the probability of (3) will be low, since we expect $\dim(\hat{W} \cap \hat{\kappa}) = |W \cap A|$. In general, vault sizes need to be quite large in order to have good estimates for security against brute force attacks. Therefore, we will assume the assumptions required to neglect the possibility of false positives hold.

6.3 Security and Considerations

We can estimate the security of this system by a similar method as for the PFV system. In particular, we want to find a parameter δ such that if $\delta < t$ elements of the vault are taken and the second coordinates of these elements are considered, then with high probability, these δ elements belong to $\hat{\kappa}$.

Lemma 6.3.1. *[46] Let $\ell \leq \delta \leq n$. The number of $\delta \times n$ matrices over $\mathbb{F}_q$ of rank ℓ is given by*

$$N_q(\ell, \delta, n) = \frac{\left(\prod_{i=0}^{\ell-1} q^n - q^i\right)\left(\prod_{i=0}^{\ell-1} q^\delta - q^i\right)}{\prod_{i=0}^{\ell-1} q^\ell - q^i}.$$

In the following, a strategy is described which tries to find a set in $\mathbb{F}_q^n$ containing ℓ linearly independent vectors that are meant to reveal the authentic features. Assume now that there are t authentic points and $r - t$ chaff points, with the set of features $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t\}$ a set of random elements of $\mathbb{F}_q^k$. We can assume that the second coordinates of the authentic set $\{\boldsymbol{x}_1\kappa, \ldots, \boldsymbol{x}_t\kappa\}$ contains a set of ℓ linearly independent vectors in $\mathbb{F}_q^n$, since from Lemma 6.3.1 the probability that $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t\}$ contains ℓ independent elements rapidly approaches 1 by Lemma 6.3.1. Now, the expected number of subsets of size δ out of $r > \delta$ random points in $\mathbb{F}_q^n$ that span a ℓ-dimensional space can be estimated by

$$\alpha_q(\ell, \delta, n) = \frac{\binom{r}{\delta} N_q(\ell, \delta, n)}{q^{\delta n}}.$$

Ideally an attacker would want to find $\delta_0 \leq |A| = t$ so that $\alpha_q(\ell, \delta_0, n) < 1$ in order to have a high probability of recovering the key in the event that the δ_0 points span a space of dimension ℓ. On the other side, to counter this type of attack, one tries to keep ℓ very close to t and r large enough, so that α_q is not too small.

We will approximate the complexity of a brute force attack following this approach. The attack is similar in approach to that proposed in [55], although adjusted for the SFV variant.

It takes $n(\delta^2 - \delta)/2$ operations to row reduce a $\delta \times n$ binary matrix. We will use this approximation, even when the field is not binary. Hence, following the proof of [55] we obtain the following upper bound for the expected time to recover the key.

Corollary 6.3.2. *In the above setting, let δ_0 be so that $\alpha_2(\ell, \delta_0, n) < 1$. On average, an attacker can recover the secret key in $C \cdot (r/t)^\ell$ operations, where $C < .55 \cdot n(\ell^2 - \ell)(r - \ell)nl$.*

Proof. Let δ_0 be such that $\alpha_q(\ell, \delta, n) \leq 1$. From [15], the average number of attempts for a user to guess ℓ points in the authentic set is

$$\frac{\binom{r}{\ell}}{\binom{t}{\ell}} < 1.1(r/t)^\ell,$$

for $r > t > 5$. We proceed as follows:

1. Choose a random set, T, of size ℓ from the second coordinates of the vault and compute the echelon form of T. This requires approximately $n(\ell^2 - \ell)/2$ operations.

2. Guess a point $\boldsymbol{v}$ from the second coordinate of the vault, but not in T. Check if $\boldsymbol{v} \in \langle T \rangle_{\mathbb{F}_q}$. It requires nl operations to check if $\boldsymbol{v} \in \langle T \rangle_{\mathbb{F}_q}$ and we must check at most $r - |T|$ points.

3. If no point is found, discard T and go to 1. If a point is found, add it to T and go to 2. Continue until $|T| = \delta_0$.

Under the assumption that the only time a point is found in step 3 occurs when T corresponds to a subset of A, the complexity can be approximated by counting the number of expected operations to choose ℓ elements of A and the approximate number of operations to rebuild the key. If it takes approximately $1.1(r/t)^\ell$ attempts to guess a valid set for T, then all previous attempts failed. For each incorrect choice of T, the process of verifying that T is incorrect requires at most

$$\frac{n(\ell^2 - \ell)}{2}(r - \ell)nl$$

operations. The result follows.

□

Remark 6.3.3. Suppose that the number of reliably extractable features is fixed by some feature extraction algorithm and suppose this number is t. Then in a PFV scheme, the minimum distance of the Reed-Solomon code underpinning the system will be $t-\ell+1$ and the number of keys will be q^ℓ for some $\ell < t$. In order to increase the number of keys while maintaining a fixed minimum distance, a designer would have to increase q. Parameter trade-offs are different in the subspace metric. As an illustration, if we choose a spread code $\mathcal{S} \subset \mathrm{Gr}_q(\ell, c\ell)$, we can take any $\ell \leq t$ and encode our features as a codeword from $\mathcal{S}$. The minimum distance will then be 2ℓ and the number of keys will be $\frac{q^{c\ell}-1}{q^\ell-1}$. Note that we can increase the number of keys and the error correction capabilities simultaneously. We can also fix ℓ and still increase the number of keys. Furthermore, these considerations do not rely on changing the field size. This means one that can trade-off storage space without compromising the parameters of the system or possibly having to modify the feature extraction algorithm to work with other fields.

Remark 6.3.4. We also would like to make a remark about the features. In order for the features for the PFV vault to be applicable, they must be representable as elements of $\mathbb{F}_q$ for some q. If $q = 2^m$, then we have the canonical isomorphism $\mathbb{F}_{2^m} \cong \mathbb{F}_2^m$. Therefore, features for using the SFV scheme are compatible with features from the PFV scheme. In the PFV scheme, the key is dependent on the representation of the features (the coefficients of the key polynomial must come from $\mathbb{F}_{2^m}$. In the SFV version, the matrix, κ, which encodes the key, $\hat{\kappa}$, must have size $m \times n$, but $\hat{\kappa}$ can have dimension smaller than or equal to m. This allows for more flexibility in designing an appropriate constant-dimension code for the system.

Appendix A

MAGMA Code

A.1 Computing Elements of Rank One

```
// Input must be a matrix G with coefficients belonging to a field F
// of cardinality q^m

RankOne := function(G)
    G := EchelonForm(G);
    G := RowSubmatrix(G, 1, Rank(G));
    k := Rank(G);
    n := Ncols(G);

    Eqn_Matrix := G;
    for i := 1 to k by 1 do
        for j := 1 to n by 1 do
            Eqn_Matrix[i,j] := Eqn_Matrix[i,j]^(q)-Eqn_Matrix[i,j];
        end for;
    end for;

    Eqn_Matrix_Exp := ZeroMatrix(BaseField(F), k, m*n);
    for i := 1 to k by 1 do
        for j := 1 to n by 1 do
            for jj := 1 to m by 1 do
                Eqn_Matrix_Exp[i, m*(jj-1) + j] := Eltseq(Eqn_Matrix[i,j])[jj];
            end for;
        end for;
    end for;
    Mq := BasisMatrix(Nullspace(Eqn_Matrix_Exp));
    M := ZeroMatrix(F, Nrows(Mq), Ncols(Mq));
    for i := 1 to Nrows(Mq) by 1 do
        for j := 1 to Ncols(Mq) by 1 do
```

```
            M[i,j] := (F ! Mq[i,j]);
        end for;
    end for;

    return M*G;
end function;
```

A.2 Generator Matrix for Gabidulin Code

```
// A random Gabidulin code is constructed. This function assumes that F is a field of
// size q^m .The parameters must satisfy k <= n <= m.

Gabidulin := function (n,k)
    gq := ZeroMatrix(BaseField(F), m, n);
    while Rank(gq) lt n do
        g := RandomMatrix(F, 1, n);
        for i := 1 to m by 1 do
            for j := 1 to n by 1 do
                gq[i,j] := Eltseq(g[1,j])[i];
            end for;
        end for;
    end while;

    G := ZeroMatrix (F, k, n);
    for i := 0 to k-1 by 1 do
        for j := 1 to n by 1 do
            G[i+1,j] := g[1, j]^(q^i);
        end for;
    end for;
    return G;
end function;
```

A.3 Cryptanalysis of CS Variant

```
// This algorithm constructs a random column scrambling matrix and proceeds to
// compute the elements of rank one in the extended matrix. G must be a generator
// matrix for the Gabidulin code of size k by n. t is the error correction capability
// of the Gabidulin code and 0 < t1 < t is a design parameter.

// Construct P;

P := ZeroMatrix(F, n, n);
```

```
while Rank(P) lt n do
    Q1 := RandomMatrix (F, n, t-t1);
    Q2q := RandomMatrix (BaseField(F), n, n-t+t1);
    Q2 := ZeroMatrix (F, n, n-t+t1);
    for i := 1 to n by 1 do
        for j := 1 to n-t+t1 by 1 do
            Q2[i,j] := (F ! Q2q[i,j]);
        end for;
    end for;
    P := HorizontalJoin (Q1, Q2);
end while;

// Create error e of rank at most t1 and the public generator matrix Gpub.
// Construct y, an encrypted message.

etemp := RandomMatrix (F, 1, t1);
Eq := RandomMatrix (BaseField(F), t1, n);
E := ZeroMatrix(F, t1, n);
for i := 1 to t1 by 1 do
    for j := 1 to n by 1 do
        E[i,j] := (F ! Eq[i,j]);
    end for;
end for;

e := etemp*E;
Gpub := Gabidulin(n,k)*P^(-1);
message := RandomMatrix(F, 1, k);
y := message*Gpub + e;

// Calculate U a Grassmann support matrix for the space spanned by the elements
// of rank one in the extended matrix and H a parity check matrix.

Gext := VerticalJoin (Gpub, e);
Gtemp := Gext;
for l := 1 to t1-1 by 1 do
    for i := 1 to k+1 by 1 do
        for j := 1 to n by 1 do
            Gtemp[i,j] := Gtemp[i,j]^(q);
        end for;
    end for;
Gext := VerticalJoin (Gext, Gtemp);
end for;
```

```
U := RankOne (Gext);
Ctemp := LinearCode(U);
H := ParityCheckMatrix(Ctemp);


// Compute the message

message := Solution(Gpub*Transpose(H),  y*Transpose(H));
```

Bibliography

[1] M. Agrawal, N. Kayal, and N. Saxena. PRIMES is in P. *Ann. of Math*, 2:781–793, 2002.

[2] R. Ahlswede, N. Cai, S.-Y.R. Li, and R.W. Yeung. Network information flow. *Information Theory, IEEE Transactions on*, 46(4):1204–1216, Jul 2000.

[3] M. Baldi, M. Bianchi, F. Chiaraluce, J. Rosenthal, and D. Schipani. Enhanced public key security for the McEliece cryptosystem. *Journal of Cryptology*, pages 1–27, 2014.

[4] A. Becker, A. Joux, A. May, and A. Meurer. How $1 + 1 = 0$ improves information set decoding. *Advances in Cryptology*, 2012.

[5] T. P. Berger. Isometries for rank distance and permutation group of Gabidulin codes. *Information Theory, IEEE Transactions on*, 49(11):3016–3019, Nov 2003.

[6] T. P. Berger and P. Loidreau. How to mask the structure of codes for a cryptographic use. pages 63–79, 2005.

[7] E. Berlekamp, R.J. McEliece, and H.C.A. Van Tilborg. On the inherent intractability of certain coding problems (corresp.). *Information Theory, IEEE Transactions on*, 24(3):384–386, May 1978.

[8] D. J. Bernstein, T. Lange, and C. Peters. Attacking and defending the McEliece cryptosystem. In Johannes Buchmann and Jintai Ding, editors, *Post-Quantum Cryptography*, volume 5299 of *Lecture Notes in Computer Science*, pages 31–46. Springer Berlin Heidelberg, 2008.

[9] A. Betten, M. Braun, H. Fripertinger, A. Kerber, A. Kohnert, and A. Wasserman. *Error-Correcting Linear codes*, volume 18. Springer-Verlag, 2006.

[10] J.F. Buss, Gudmund G.S., and O.S. Jeffrey. The computational complexity of some problems of linear algebra. In Rüdiger Reischuk and Michel Morvan, editors, *STACS 97*, volume 1200 of *Lecture Notes in Computer Science*, pages 451–462. Springer Berlin Heidelberg, 1997.

[11] A. Canteaut and H. Chabanne. *A Further Improvement of the Work Factor in an Attempt at Breaking McEliece's Cryptosystem.* Rapports de recherche. Institut national de recherche en informatique et en automatique, 1994.

[12] F. Chabaud and J. Stern. The cryptographic security of the syndrome decoding problem for rank distance codes. In *Advances in Cryptology - ASIACRYPT '96, International Conference on the Theory and Applications of Cryptology and Information Security, Kyongju, Korea, November 3-7, 1996, Proceedings*, pages 368–381, 1996.

[13] D. Cheung, D. Maslov, J. Mathew, and D.K. Pradhan. On the design and optimization of a quantum polynomial-time attack on elliptic curve cryptography. In Yasuhito Kawano and Michele Mosca, editors, *Theory of Quantum Computation, Communication, and Cryptography*, volume 5106 of *Lecture Notes in Computer Science*, pages 96–104. Springer Berlin Heidelberg, 2008.

[14] W.Y. Choi, S.B. Pan, Kim J.-M., Chung Y., and D. won Hong. Fast polynomial reconstruction attack against fuzzy fingerprint vault. In *Information Science and Service Science (NISS), 2011 5th International Conference on New Trends in*, volume 2, pages 299–302, Oct 2011.

[15] C. Clancy. Secure smartcard-based fingerprint authentication. In *ACM Workshop on Biometrics: Methods and Applications*, pages 45–52, 2003.

[16] A. Cossidente, G. Marino, and F. Pavese. Non-linear maximum rank distance codes. *Designs, Codes and Cryptography*, pages 1–13, 2015.

[17] J. de la Cruz, M. Kiermaier, A. Wassermann, and W. Willems. Algebraic structures of MRD codes. arXiv:1502.02711, 2015.

[18] P. Delsarte. The association schemes of coding theory. In Jr. Hall, M. and J.H. van Lint, editors, *Combinatorics*, volume 16 of *NATO Advanced Study Institutes Series*, pages 143–161. Springer Netherlands, 1975.

[19] A.G. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran. Network coding for distributed storage systems. *Information Theory, IEEE Transactions on*, 56(9):4539–4551, September 2010.

[20] H. Dinh, C. H. Moore, and A. Russell. McEliece and Niederreiter cryptosystems that resist quantum fourier sampling attacks. In Phillip Rogaway, editor, *Advances in Cryptology – CRYPTO 2011*, volume 6841 of *Lecture Notes in Computer Science*, pages 761–779. Springer Berlin Heidelberg, 2011.

[21] Bernstein D.J., T. Lange, C. Peters, and H.C.A. van Tilborg. Explicit bounds for generic decoding algorithms for code-based cryptography. *Pre-proceedings of WCC 2009*, page 168–180, 2009.

[22] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM J. Comput.*, 38(1):97–139, March 2008.

[23] T. Etzion. Problems on q-analogs in coding theory. *ArXiv*, 1305.6126, 2013.

[24] T. Etzion and N. Silberstein. Error-correcting codes in projective spaces via rank-metric codes and Ferrers diagrams. *IEEE Trans. Inform. Theory*, 55(7):2909–2919, 2009.

[25] M. Finiasz and N. Sendrier. Security bounds for the design of code-based cryptosystems. In Mitsuru Matsui, editor, *Advances in Cryptology – ASIACRYPT 2009*, volume 5912 of *Lecture Notes in Computer Science*, pages 88–105. Springer Berlin Heidelberg, 2009.

[26] E. M. Gabidulin. Theory of codes with maximum rank distance. *Problemy Peredachi Informatsii*, 21(1):3–16, 1985.

[27] E. M. Gabidulin, A. V. Ourivski, B. Honary, and B. Ammar. Reducible rank codes and their applications to cryptography. *IEEE Transactions on Information Theory*, 49(12):3289–3293, 2003.

[28] E. M. Gabidulin, A. V. Paramonov, and O. V. Tretjakov. Ideals over a non-commutative ring and their application in cryptology. In *Proceedings of the 10th Annual International Conference on Theory and Application of Cryptographic Techniques*, EUROCRYPT'91, pages 482–489, Berlin, Heidelberg, 1991. Springer-Verlag.

[29] E. M. Gabidulin, H. Rashwan, and B. Honary. On improving security of GPT cryptosystems. In *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pages 1110–1114, June 2009.

[30] P. Gaborit, O. Ruatta, and J. Schrek. On the complexity of the rank syndrome decoding problem. *Information Theory, IEEE Transactions on*, 62(2):1006–1019, Feb 2016.

[31] P. Gaborit, O. Ruatta, J. Schrek, and G. Zémor. New results for rank-based cryptography. In *Progress in Cryptology - AFRICACRYPT 2014 - 7th International Conference on Cryptology in Africa, Marrakesh, Morocco, May 28-30, 2014. Proceedings*, pages 1–12, 2014.

[32] M. Gadouleau and Zhiyuan Y. Constant-rank codes and their connection to constant-dimension codes. *Information Theory, IEEE Transactions on*, 56(7):3207–3216, July 2010.

[33] M. Gadouleau and Z. Yan. Packing and covering properties of rank metric codes. *IEEE Trans. Info. Theory*, pages 3873–3883, 2008.

[34] J. K. Gibson. Severely denting the Gabidulin version of the McEliece public key cryptosystem. *Des. Codes Cryptography*, 6(1):37–45, July 1995.

[35] M. Giorgetti and A. Previtali. Galois invariance, trace codes and subfield subcodes. *Finite Fields and Their Applications*, 16(2):96–99, 2010.

[36] E. Gorla and A. Ravagnani. Partial spreads in random network coding. *Finite Fields Appl.*, 26:104–115, March 2014.

[37] V. Guruswami. *List Decoding of Error-Correcting Codes*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[38] S. Han, S. Liu, K. Chen, and D. Gu. Proofs of data possession and retrievability based on MRD codes, 2013.

[39] A-L. Horlemann-Trautmann and K. Marshall. New criteria for MRD and Gabidulin codes and some rank-metric code constructions. *arXiv:1507.08641*, 2015.

[40] A.L. Horlemann-Trautmann, K. Marshall, and J. Rosenthal. Extension of Overbeck's attack for Gabidulin based cryptosystems. *ArXiv*, 1511.01549, 2015.

[41] A. Juels and M. Wattenberg. A fuzzy commitment scheme. In *Proc. 6th ACM Conference on Computer and Communications Security*, CCS '99, pages 28–36, New York, NY, USA, 1999. ACM.

[42] M. Juels, A.and Sudan. A fuzzy vault scheme. *Des. Codes Cryptography*, 38(2):237–257, February 2006.

[43] R. Jurrius and R. Pellikaan. On defining generalized rank weights. *arXiv*, 1506.02865, 2015.

[44] R. Koetter and F. R. Kschischang. Coding for errors and erasures in random network coding. *Information Theory, IEEE Transactions on*, 54(8):3579–3591, August 2008.

[45] A. Kshevetskiy and E. Gabidulin. The new construction of rank codes. In *Proceedings of the International Symposium on Information Theory (ISIT) 2005*, pages 2105–2108, Sept 2005.

[46] D. Laksov and A. Thorup. Counting matrices with coordinates in finite fields and of fixed rank. *Mathematica Scandinavica*, 74:19–33, 1994.

[47] A. K. Lenstra, H. W. Lenstra, Jr., M. S. Manasse, and J. M. Pollard. The number field sieve. In *Proceedings of the Twenty-second Annual ACM Symposium on Theory of Computing*, STOC '90, pages 564–572, New York, NY, USA, 1990. ACM.

[48] R. Lidl and H. Niederreiter. *Introduction to Finite Fields and their Applications*. Cambridge University Press, Cambridge, London, 1986. 1994 Revised edition.

[49] P. Loidreau. Designing a rank metric based McEliece cryptosystem. In *Proceedings of the Third International Conference on Post-Quantum Cryptography*, PQCrypto'10, pages 142–152, Berlin, Heidelberg, 2010. Springer-Verlag.

[50] G. Lunardon, R. Trombetti, and Y. Zhou. Generalized twisted Gabidulin codes. *ArXiv:1507.07855*, July 2015.

[51] Baldi. M., M. Bianchi, F. Chiaraluce, J. Rosenthal, and D. Schipani. On fuzzy syndrome hashing with LDPC coding. In *Proc. 4th Int. Symp. Applied Sciences in Biomedical and Communication Technologies (ISABEL)*, pages 1–5. ACM, 2011.

[52] F. Manganiello, E. Gorla, and J. Rosenthal. Spread codes and spread decoding in network coding. In *Proc. IEEE Int. Symp. Information Theory*, pages 881–885, 2008.

[53] K. Marshall, D. Schipani, A-L. Trautmann, and J. Rosenthal. Subspace fuzzy vault. In Marco Baldi and Stefano Tomasin, editors, *Physical and Data-Link Security Techniques for Future Communication Systems*, volume 358 of *Lecture Notes in Electrical Engineering*, pages 163–172. Springer International Publishing, 2016.

[54] R. McEliece. *The Theory of Information and Coding*, volume 86. Cambridge University Press, 2004.

[55] P. Mihailescu, A. Munk, and B. Tams. The fuzzy vault for fingerprints is vulnerable to brute force attack. In *Proc. Biometric Special Interests Group*, pages 43–54, 2009.

[56] E.H. Moore. A two-fold generalization of Fermat's theorem. page 189–199, 1896.

[57] K. Morrison. Equivalence for rank-metric and matrix codes and automorphism groups of Gabidulin codes. *IEEE Transactions on Information Theory*, Vol 60(11):1–12, 2014.

[58] K. Nandakumar, A.K. Jain, and S. Pankanti. Fingerprint-based fuzzy vault: Implementation and performance. *Information Forensics and Security, IEEE Transactions on*, 2(4):744–757, Dec 2007.

[59] O. Ore. On a special class of polynomials. *Trans. American Math. Soc.*, 35:559–584, 1933.

[60] A. V. Ourivski and E. M. Gabidulin. Column scrambler for the GPT cryptosystem. *Discrete Appl. Math.*, 128(1):207–221, May 2003.

[61] A. V. Ourivski and T. Johansson. New technique for decoding codes in the rank metric and its cryptography applications. *Probl. Inf. Transm.*, 38(3):237–246, July 2002.

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

[63] E. Prange. The use of information sets in decoding cyclic codes. *Information Theory, IRE Transactions on*, 8(5):5–9, September 1962.

[64] H. Rashwan, E.M. Gabidulin, and B. Honary. A smart approach for GPT cryptosystem based on rank codes. In *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pages 2463–2467, June 2010.

[65] A. Ravagnani. Rank-metric codes and their duality theory. *Designs, Codes and Cryptography*, pages 1–20, 2015.

[66] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM*, 21(2):120–126, February 1978.

[67] G. Robert, P. Loidreau, and D. Augot. Rank metric and Gabidulin codes in characteristic zero. *arXiv:1305.4047*, 2013.

[68] J. Rosenthal, N. Silberstein, and A.-L. Trautmann. On the geometry of balls in the Grassmannian and list decoding of lifted Gabidulin codes. *Designs, Codes and Cryptography*, 73:393–416, 2014.

[69] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27, 1948.

[70] J. Sheekey. A new family of linear maximum rank distance codes. *ArXiv:1504.01581*, April 2015.

[71] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, October 1997.

[72] V. M. Sidelnikov and S. O. Shestakov. On insecurity of cryptosystems based on generalized Reed-Solomon codes. *Discrete Mathematics and Applications*, 2:439–444, 1992.

[73] N. Suresh Babu. Studies on rank distance codes. *Ph. D Dissertation*, 1995.

[74] A.-L. Trautmann. A lower bound for constant dimension codes from multi-component lifted MRD codes. *arXiv:1301.1918 [cs.IT]*, 2013.

[75] H. C. A. van Tilborg. *Encyclopedia of Cryptography and Security*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[76] A. Wachter-Zeh. Bounds on list decoding of rank-metric codes. *IEEE Trans. Inf. Theor.*, 59(11):7268–7277, 2013.

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

[78] X. Zhang, R. Shi, and J. Ritcey. On the implementation of modified fuzzy vault for biometric encryption. In *Information Theory and Applications Workshop (ITA), 2012*, pages 56–61, Feb 2012.