On the dimension of the subfield subcodes of Hermitian codes

outline of the Ph. D. Thesis

by

Sabira El Khalfaoui

Thesis advisor:
Prof. Dr. Gábor P. Nagy

Doctoral School of Mathematics and Computer Science, Bolyai Institute
University of Szeged, Faculty of Science and Informatics
Szeged, 2020

Contents

1 Introduction

This dissertation seeks to study some classes of subfield subcodes of Hermitian codes. These papers [EKN19; EKN20] present the results of our research that aims at revealing the properties and the structure of the underlying classes of codes. Furthermore, we intend to examine the potential of this class of codes to improve the practicality of the McEliece cryptosystem. The problems we treat belong to coding theory and their applications to cryptography. They have a common aspect, which is security that refers to code-based cryptography. Here, we briefly introduce the preliminaries and the topics with a short history that describes the main results.

The result of the paper [EKN19] is discussed in chapter 3 which is about the proof of the true dimension of Hermitian subfield subcodes for specific parameters. Finding the true dimension of the subfield subcodes of linear codes was studied by many researchers who tried to improve the general bound of the dimension to obtain a code with a large dimension and minimum distance. We only present this problem for the class of Goppa codes in chapter 2. The solution to this problem allows us to find out more facts about the class of codes and which can later lead to further research.

In chapter 4, we rely on the paper [EKN20] which deals with the problem of approximating the true dimension of subfield subcodes of Hermitian codes by an explicit formula. We describe the statistical set up to tackle the experimental study to analyze the datasets of the true dimension of different subfield subcodes of Hermitian codes. The datasets were computed using our GAP package `HERmitian` [NEK19]. Based on adjusting the distribution to the underlying datasets using the method `fitmethis` of MATLAB [TM19; Cas20], we found that the extreme value distribution is the most suitable one.

Chapter 5 is dedicated to applying subfield subcodes of Hermitian codes in cryptography in which we precisely suggest the mentioned class of codes for McEliece cryptosystem. Mainly, we give a formula of the public key size in terms of the code rate using the result of the paper [EKN20], see also chapter 4. We describe an overview of post-quantum cryptography in which code-based cryptography is part of, representing the central area of applications concerning coding theory. This overview shows the importance of designing cryptographic schemes that can resist post-quantum attacks since the presence of quantum computer threatens the so-called classical cryptography. All cryptosystems are based on a computationally hard problem such as integer factorization (RSA), or discrete logarithm problem (ECC, El Gamal).

Acknowledgment

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Gábor P. Nagy for his patience, motivation, and knowledge which supported my Ph.D. study and the related research. His guidance helped me to learn many things in different areas of research. His advice was invaluable for my research and the writing of this thesis.

My sincere thanks also go to both Bolyai Institute and Stipendium Hungaricum Foundation. They provided me an opportunity to join Ph.D. studies, and they gave access to all research facilities. Without their precious support, it would not be possible to conduct this research.

Large part of the results were obtained in the framework of the research project *Security Enhancing Technologies for the Internet of Things (SETIT)* [1]. The motivating discussions and feedback helped a lot for carrying out the presented work.

Last but not least, I would like to thank my family: my grandparents, my parents, my siblings, and my friends for supporting me spiritually throughout writing this thesis and my life in general.

2 Preliminaries

In the last decades, there has been a huge need for reliable digital data transmission and storage systems. This need has grown thanks to the appearance of high-speed data networks for the interchange, the treatment and the storage of digital information in

[1] Project no. 2018-1.2.1-NKP-2018-00004 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2018-1.2.1-NKP funding scheme.

both the public and private sectors. It is necessary to incorporate communications and computer technology to design such systems. Obtaining a reliable data reproduction can be done by controlling the occurred errors which is the major aims of a designer.

In 1948, Shannon introduced a mathematical framework to describe communication channels with or without errors. In his famous paper [LC01], Shannon demonstrated the existence of encoding and decoding schemes. This work was to some extent inspired by Ludwig Boltzmann's work in statistical physics. Hamming gave the idea of detecting and correcting errors. It was a consequence of resolving the problem when his computer came to turn off every time it detected an error. Shannon's Second Theorem concerns channel coding. In other words, it adds extra information to a message that is intended to be sent in a noisy channel which protects it against transmission errors. Moreover, this extra information permits us to detect or even correct some transmission errors which gave birth to error-correcting codes theory.

2.1 Subfield subcodes of linear codes

Definition 2.1. *Let C be a $[n,k]$ linear code over $\mathbb{F}_q$, where $q = r^m$ is a prime power. The $\mathbb{F}_q/\mathbb{F}_r$ subfield subcode $C|_{\mathbb{F}_r}$ of C is by definition the set*

$$C|_{\mathbb{F}_r} = C \cap \mathbb{F}_r^n$$

of all codewords in C with components in $\mathbb{F}_r$.

The $\mathbb{F}_q/\mathbb{F}_r$ subfield subcode is a linear (n, k_0, d_0) code with $d \leq d_0 \leq n$ and $n - k \leq n - k_0 \leq m(n-k)$. A parity check matrix of C over $\mathbb{F}_q$ yields at most $m(n-k)$ linearly independent parity equations over $\mathbb{F}_r$ for the subfield subcodes $C|_{\mathbb{F}_r}$.

In general, the minimum distance of the subfield subcode is bigger than the minimum distance of the original one.

Let $T_{\mathbb{F}_q/\mathbb{F}_r}$ be the trace polynomial in the field $\mathbb{F}_q$ with respect to $\mathbb{F}_r$, that is

$$T_{\mathbb{F}_q/\mathbb{F}_r}(x) = x + x^r + ... + x^{r^{m-1}}.$$

For a vector $c \in \mathbb{F}_q^n$, $T_{\mathbb{F}_q/\mathbb{F}_r}(c) = (T_{\mathbb{F}_q/\mathbb{F}_r}(c_1), \cdots, T_{\mathbb{F}_q/\mathbb{F}_r}(n))$. For a linear code C of length n and dimension k over $\mathbb{F}_q$, $T_{\mathbb{F}_q/\mathbb{F}_r}(C)$ is a linear code with the same length of C and dimension k_1 over $\mathbb{F}_r$.

Delsarte has come up with a very important result which relates the subfield subcode to the trace code in the following theorem:

Theorem 2.1 ([Del75])**.** *Let C be a $[n,k]$ linear code over $\mathbb{F}_q$. Then $(C|_{\mathbb{F}_r})^{\perp} = T_{\mathbb{F}_q/\mathbb{F}_r}(C^{\perp})$ holds.*

The class of subfield subcodes and trace codes held the attention of many researchers. A lot of work was done on the class of subfield subcodes by Stichtenoth [Sti90], and it was improved upon in [SMS97]. The study of trace codes was made by Van der Vlugt [Vlu91; VDV91]. Roseiro stated the relation between trace codes and Goppa codes which was established in [Ros+92] using the tool given by Delsarte (see [Del75]).

Lemma 2.2. *Let C be an $[n, K]$ linear codes over the finite field $\mathbb{F}_q$, where $q = r^m$. The subfield subcode of C satisfies:*

$$\dim_{\mathbb{F}_r}(C \cap \mathbb{F}_r^n) = n - m(n - K) + \dim_{\mathbb{F}_r}(\ker(T_{\mathbb{F}_q/\mathbb{F}_r})). \tag{1}$$

2.2 The true dimension of binary Goppa codes

By choosing the parameters of the binary Goppa code in an appropriate way, it is possible to increase its dimension and minimum distance. In this section, we investigate the problem of finding the true dimension of Goppa codes which is considered as subfield subcodes of generalized Reed-Solomon codes. We summarize two strategies that are used to get new bound for the dimension of Goppa codes which was the aim of many researchers [Ros+92; BS95; Véro01; Vér05; Vér98].

The first strategy is about the link between the parity check matrix $\tilde{H}$ of Goppa codes and the parity check matrix H of **GRS** codes. The parity check matrix defined above does not generate the dual of Goppa codes because it is defined over $\mathbb{F}_q$. We can compute the parity check matrix $\tilde{H}$ over $\mathbb{F}_r$ from the parity check matrix H over $\mathbb{F}_q$ by converting each column vector of H to a column vector over $\mathbb{F}_r$. Therefore, computing the dimension of $\Gamma\,(L, g)$ is equivalent to computing the rank of $\tilde{H}$. H has $\deg g(z)$ rows, then $\tilde{H}$ has $m \deg g(z)$ rows which are not necessarily independent. This strategy has been stated in [Vér98], where the author explained (with an example [Vér05]) how to improve the bound $k \geq n - m \deg g(z)$, by looking for some polynomials and choosing a special basis, when computing $\tilde{H}$ from H, in order to find linear dependent rows [Vér05].

The second strategy used Delsarte's result [Del75] so as to define codes with large dimension k. It is based on using the image of the dual code under the trace map with a rank that is equal to redundancy [Ros+92]. The objective of the strategy is to find polynomials $g(z)$ such that the trace map has a large kernel. This strategy was the main idea of [Ros+92], it was applied to the classes of primitive binary Goppa codes whose polynomial satisfies $G^{2^s}(X) \equiv G(X)\,(\mod X^{2^{2s}} + X)$.

3 Algebraic geometry codes

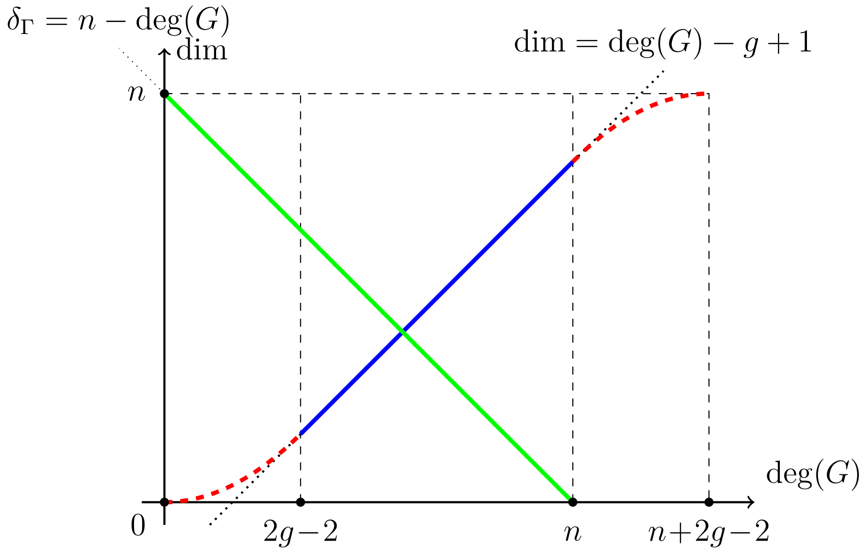
3.1 Algebraic geometry codes (AG codes)

Let q be a prime power, and $\mathbb{F}_q$ be the finite field of order q. Let $\mathcal{X}$ be an algebraic curve, i.e., an affine or projective variety of dimension one, which is absolutely irreducible and nonsingular and whose defining equations are (homogeneous) polynomials with coefficients in $\mathbb{F}_q$. Let g be the genus of $\mathcal{X}$. In the following, $P_1, \cdots, P_n$ are pairwise distinct places on $\mathcal{X}$ and D is the divisor $D = P_1 + ... + P_n$. Furthermore, G is another divisor with support disjoint from D.

Definition 3.1. *The algebraic geometry code $C_{\mathscr{L}}(D, G)$ associated with the divisors D and G is defined as*

$$C_{\mathscr{L}}(D, G) = \{(f(P_1), f(P_2), ..., f(P_n)) \mid f \in \mathscr{L}(G)\} \subseteq \mathbb{F}_q^n.$$

Figure 1: Dimension and designed minimum distance of AG codes



In other words, $C_{\mathscr{L}}(D, G)$ is the image of $\mathscr{L}(G)$ under the evaluation map

$$\mathscr{L}(G) \ni f \mapsto (f(P_1), ..., f(P_n)) \in \mathbb{F}_q^n.$$

Theorem 3.1 ([Sti09])**.** *$C_{\mathscr{L}}(D, G)$ is a $[n, k, d]$ codes with parameters:*

- *$k = \ell(G) - \ell(G - D)$ where $\ell(G) = \dim \mathscr{L}(G)$*
- *$d \geq n - \deg G$*

We illustrate the behavior of the dimension k of $C_{\mathscr{L}}(D, G)$ depending on the degree of the divisor G by Figure 1. In fact, Theorem 3.1 implies the exact value $k = \deg(G) - g + 1$ provided $2g - 2 < \deg(G) < n$. Furthermore, if $\deg(G) > n + 2g - 2$, then $k = n$. In the intervals $[0, 2g - 2]$, and $[n, n + 2g - 2]$, the dimension depends on the specific structure of the divisor G.

3.2 Hermitian codes

An important class of AG codes that have good properties is the class of Hermitian codes. This class is constructed by employing Hermitian curves over a finite field. The Hermitian curve $\mathscr{H}_q$ over $\mathbb{F}_{q^2}$ in affine coordinates has the form

$$\mathscr{H}_q : Y^q + Y = X^{q+1}.$$

Its rational points are points of the projective plane $PG(2, q^2)$, satisfying the homogenous equation $Y^qZ + YZ^q = X^{q+1}$. It is easy to verify that $\mathscr{H}_q$ is nonsingular, then its genus is $g = q(q-1)/2$ by the genus formula. With respect to the line $Z = 0$ at infinity, $\mathscr{H}_q$ has one infinite point $P_\infty = (0:1:0)$ and q^3 affine rational points $P_1, \ldots, P_{q^3}$, which make the class of Hermitian curves interesting since they attain the maximal number of rational points for the famous Hasse-Weil bound [Men+13]. As usual, we also look at the curve $\mathscr{H}_q$ as the smooth curve defined over the algebraic closure $\bar{\mathbb{F}}_{q^2}$. Then, there is a one-to-one correspondence between the points of $\mathscr{H}_q$ and the places of the function field $\bar{\mathbb{F}}_{q^2}(\mathscr{H}_q)$.

With a Hermitian code we mean a functional AG code of the form $C_{\mathscr{L}}(D, G)$, where the divisor D is defined as the sum $P_1 + \cdots + P_{q^3}$ of all affine rational points of $\mathscr{H}_q$. In our investigations, the divisor G can take two forms. In the *1-point case,* we set $G = sP_\infty$ with integer s. In the *degree 3 case,* we put $G = sP$, where P is a place of degree 3. Let P_1, P_2, P_3 be the extensions of P in the constant field extension of $\mathbb{F}_{q^2}(\mathscr{H}_q)$ of degree 3. Then P_1, P_2, P_3 are degree one places of $\mathbb{F}_{q^6}(\mathscr{H}_q)$ and, up to labeling the indices, $P_{j+1} = \mathrm{Frob}(P_j)$ where Frob is the q^2-th Frobenius map and the indices are taken modulo 3. Also, P may be identified with the $\mathbb{F}_{q^2}$-rational divisor $P_1 + P_2 + P_3$ of $\mathbb{F}_{q^6}(\mathscr{H}_q)$. Functional AG codes of the form $C_{\mathscr{L}}(D, sP_\infty)$ and $C_{\mathscr{L}}(D, sP)$ will be called 1-point Hermitian codes, and Hermitian codes over a degree 3 place, respectively. In the 1-point case, the basis of the Riemann-Roch space $\mathscr{L}(sP_\infty)$ can be given explicitly by [Ste12]:

$$\mathcal{M}(s) := \left\{ x^i y^j \mid 0 \le i \le q^2 - 1, 0 \le j \le q - 1, qi + (q+1)j \le s \right\}.$$

In the degree 3 case, the Riemann-Roch space

$$\mathscr{L}(sP) = \left\{ \frac{f}{(\ell_1 \ell_2 \ell_3)^u} \mid f \in \mathbb{F}_{q^2}[X, Y], \deg f \le 3u, v_{P_i}(f) \ge v \right\} \cup \{0\}.$$

can be computed, see [KN13]. In this formula, $\ell_i = 0$ is the equation of the tangent line of $\mathscr{H}_q$ at P_i, and $s = u(q+1) - v$, $0 \le v \le q$.

The group $\mathrm{Aut}(\mathscr{H}_q)$ of all automorphisms of $\mathscr{H}_q$ is defined over $\mathbb{F}_{q^2}$. It is a group of projective linear transformations of $PG(2, q^2)$, isomorphic to the projective unitary group $PGU(3, q)$. Furthermore, $\mathrm{Aut}(\mathscr{H}_q)$ acts doubly transitively on the set $\{P_\infty, P_1, \ldots, P_{q^3}\}$ of $\mathbb{F}_{q^2}$-rational points. As it was pointed out in [KN13], the automorphism group of $\mathscr{H}_q$ acts transitively on the set of degree 3 places of $\mathbb{F}_{q^2}(\mathscr{H}_q)$, as well. Hence, the geometry of a degree 3 place is independent on the choice of P. However, the stabilizer G_P of P in $\mathrm{Aut}(\mathscr{H}_q)$ is not transitive on the set of $q^3 + 1$ rational points. In fact, G_P is a cyclic group of order $q^2 - q + 1$ and the number of G_P-orbits on the set of rational points is $q + 1$. (See [CKT99; KN13], where [CKT99, Section 4.2] holds for any characteristic.)

3.3 On the true dimension of the subfield subcodes of 1–point Hermitian codes

The 1–point Hermitian code $\mathcal{H}(q^2, s)$ has length $n = q^3$, if $2g - 2 < s < n$ then the dimension is $k = s - g + 1$ and the minimum distance is $d = q^3 - s$.

Definition 3.2. *Let $\mathcal{H}(q^2, s)$ be a 1–point Hermitian code, the subfield subcode of $\mathcal{H}(q^2, s)$ is*

$$C_{q,r}(s) = \mathcal{H}(q^2, s)|_{\mathbb{F}_r}.$$

In [PJ14], the authors present an algorithm to compute $\dim C_{q,r}(s)$. Using this algorithm, the dimension of $C_{4,2}(s)$ is determined for each $s = 0, \ldots, 71$.

In [Vlu91, Proposition 3.2], the author shows

$$\dim T_{\mathbb{F}_{q^2}/\mathbb{F}_r}(\mathcal{H}(q^2, q)) = 2m + 1,$$

where $q = 2^m$. In our notation, this means

$$\dim C_{q,r}(q^3 + q^2 - 2q - 2) = q^3 - (2m + 1).$$

In particular, $\dim C_{4,2}(70) = 59$, which is confirmed by [PJ14, Table 2]. In the same table, we find $\dim C_{4,2}(s) = 1$ for $s = 0, \ldots, 31$ and $\dim C_{4,2}(32) = 5$. In the next section, we prove a formula which implies these dimensions. The following is the main result of [EKN19]

Theorem 3.2. *Let $C_{q,r}(s)$ be a subfield subcode of the Hermitian code $\mathcal{H}(q^2, s)$, $q = r^m$ is a prime power. Then*

$$\dim C_{q,r}(s) = \begin{cases} 1 & \text{for } s < \frac{q^3}{r} \\ 2m + 1 & \text{for } s = \frac{q^3}{r} \end{cases}$$

4 Estimating the dimension of Hermitian subfield subcodes

In this chapter, we study the possibility of the application of subfield subcodes of Hermitian codes in the McEliece scheme. More precisely, we do the first step by investigating the true dimension of these codes for a broad spectrum of parameters, for partial results, see [EKN19; PJ14]. Our main observation is that the true dimension of subfield subcodes of Hermitian codes can be estimated by the extreme value distribution function.

We established an approximating formula of the true dimension of the subfield subcodes of Hermitian codes. We conducted an experimental study to analyze the datasets of the true dimension of different subfield subcodes of Hermitian codes. This analysis helped us to derive new properties of their structure and led to an approach that might be useful for further research and applications. Before we tackle our contribution, we need to describe the set up of statistical formulas such as moment and expectation by mean of the extended rate function of the underlying classes of subfield subcodes of Hermitian codes.

4.1 Moments of the extended rate of subfield subcodes

In order to make our notation consistent, we make the following conventions. Let $\mathcal{X}$ be an algebraic curve over $\mathbb{F}_q$ and D, G divisors such that the AG code $C_L(D, G)$ is well defined. Assume that the objects δ and γ determine the curve $\mathcal{X}$ and the divisors D, G in a unique way. Let s be an integer and $\mathbb{F}_r$ be a subfield of $\mathbb{F}_q$. Then,

$$C_{\delta,r}^{\gamma}(s) = C_L(D, sG)|_{\mathbb{F}_r}$$

denotes the $\mathbb{F}_q/\mathbb{F}_r$ subfield subcode of the AG code $C_L(D, sG)$. The length of $C_{\delta,r}^{\gamma}(s)$ is $n = \deg(D)$.

For the integer s, let

$$R(s) = R_{\delta,r}^{\gamma}(s) = \frac{\dim_{\mathbb{F}_r} C_{\delta,r}^{\gamma}(s)}{n}$$

denote the rate of the subfield subcode $C_{\delta,r}^{\gamma}(s)$. We extend $R_{\delta,r}^{\gamma}$ to $\mathbb{R}$ in the usual way: $R_{\delta,r}^{\gamma}(x) = R_{\delta,r}^{\gamma}(\lfloor x \rfloor)$.

We can consider $R(x) = R_{\delta,r}^{\gamma}(x)$ as the distribution function of some random variable ξ. In this way, $R_{\delta,r}^{\gamma}(s)$ has an expectation $\mathsf{E}_{\delta,r}^{\gamma}$, a variance $\mathsf{Var}_{\delta,r}^{\gamma}$ and a standard deviation $\mathsf{D}_{\delta,r}^{\gamma}$. These constants can be computed from the true dimensions of the subfield subcodes.

4.2 Computed true dimensions of Hermitian subfield subcodes

Let q be a prime power. We say that the object $\delta = q$ determines the Hermitian curve $\mathscr{H}_q$ over $\mathbb{F}_{q^2}$, together with the divisor D which is the sum of affine rational points of $\mathscr{H}_q$. The objects $\gamma = \text{1-pt}$ or $\gamma = \text{deg-3}$ determine the divisor G to be equal either to the rational infinite place P_∞, or the degree 3 Hermitian place P, respectively. That being said, for any integer s and subfield $\mathbb{F}_r$ of $\mathbb{F}_{q^2}$, the Hermitian subfield subcodes

$$C_{q,r}^{\text{1-pt}}(s) = C_L(D, sP_\infty)|_{\mathbb{F}_r}, \qquad C_{q,r}^{\text{deg-3}}(s) = C_L(D, sP)|_{\mathbb{F}_r}$$

are well defined and consistent with the notation of section 4.1. In chapter 3, we denoted $C_{q,r}(s)$ by $C_{q,r}^{\text{1-pt}}(s)$. All these codes are $\mathbb{F}_r$-linear codes of length $n = q^3$.

Let $R_{q,r}^{\text{1-pt}}(s)$ and $R_{q,r}^{\text{deg-3}}(s)$ be the true rates of the codes $C_{q,r}^{\text{1-pt}}(s)$ and $C_{q,r}^{\text{deg-3}}(s)$. Using the GAP [Gap] package `HERmitian` [NEK19], we have been able to compute the true dimension values of the codes $C_{q,q}^{\text{1-pt}}(s)$, $C_{q,q}^{\text{deg-3}}(s)$ for

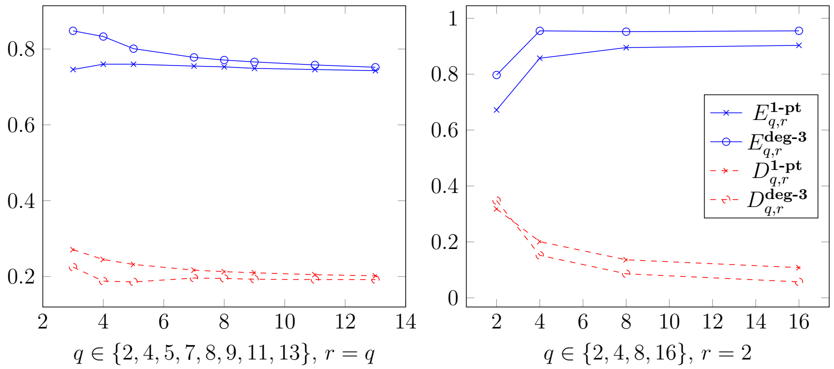
$$q \in \{2, 3, 4, 5, 7, 8, 9, 11, 13\}$$

and the binary codes $C_{q,2}^{\text{1-pt}}(s)$, $C_{q,2}^{\text{deg-3}}(s)$ for

$$q \in \{2, 4, 8, 16\}.$$

We computed the expectations $\mathsf{E}_{q,q}^{\text{1-pt}}$, $\mathsf{E}_{q,2}^{\text{1-pt}}$, $\mathsf{E}_{q,q}^{\text{deg-3}}$, $\mathsf{E}_{q,2}^{\text{deg-3}}$, the variances $\mathsf{Var}_{q,q}^{\text{1-pt}}$, $\mathsf{Var}_{q,2}^{\text{1-pt}}$, $\mathsf{Var}_{q,q}^{\text{deg-3}}$, $\mathsf{Var}_{q,2}^{\text{deg-3}}$, and the standard deviations $\mathsf{D}_{q,r}^{\mathbf{1\text{-}pt}}$, $\mathsf{D}_{q,2}^{\text{1-pt}}$, $\mathsf{D}_{q,q}^{\text{deg-3}}$, $\mathsf{D}_{q,2}^{\text{deg-3}}$ for these true rates. In Figure 2, we present the ratios $\mathsf{E}_{q,r}^{\gamma} \deg(G)/n$ and $\mathsf{D}_{q,r}^{\gamma} \deg(G)/n$, where $\gamma \in \{\text{1-pt}, \text{deg-3}\}$. While our data sets are small, these figures motivate the following open problem.

Figure 2: The ratios of expectations and standard deviations to $n/\deg(G)$



Problem 4.1. Are there constants $c_1, c_2 > 0$ such that

$$\mathsf{E}^{\text{1-pt}}_{q,q} \approx \mathsf{E}^{\text{deg-3}}_{q,q} \approx c_1 q^3/\deg(G), \qquad \mathsf{D}^{\text{1-pt}}_{q,q} \approx \mathsf{D}^{\text{deg-3}}_{q,q} \approx c_2 q^3/\deg(G),$$

where $a \approx b$ means $a/b \to 1$ with $q \to \infty$.

Our data suggests that for small q, the choice $c_1 = 0.75$ and $c_2 = 0.2$ is sound.

4.3 Distribution fitting

In general, no explicit formula is known for the true dimension of subfield subcodes of AG codes. We study the behavior of the subfield subcodes of Hermitian codes using distribution fitting methods. With fixed q, r and $\gamma \in \{\text{1-pt}, \text{deg-3}\}$ the dimensions of the subfield subcodes are given by the extended rate functions

$$R^{\text{1-pt}}_{q,q}(x), \quad R^{\text{1-pt}}_{q,2}(x), \quad R^{\text{deg-3}}_{q,q}(x), \quad R^{\text{deg-3}}_{q,2}(x).$$

Our goal is to consider these functions as distribution functions and fit some well known probability distribution functions to our experimental rate function $R(x)$.

We obtain numerical results by using the distribution fitting methods offered by MATLAB's Statistics and Machine Learning Toolbox [TM19]. To compare different distributions for a given data set, one can use the log-likelihood values for a ranking. This is implemented MATLAB's `fitmethis` function [Cas20]. We restricted ourselves to parametric distributions having at most two parameters, that is, we used `fitmethis` to compare the log-likelihood values of the following distributions: normal, exponential, gamma, logistic, uniform, extreme value, Rayleigh, beta, Nakagami, Rician, inverse Gaussian, Birnbaum-Saunders, log-logistic, log-normal and Weibull. We can summarize the results as follows:

Proposition 4.1. *1. The best fitting distribution was the extreme value distribution for $R^{\text{1-pt}}_{q,q}(x)$, $q \in \{4, 5, 7, 8, 9, 11, 13\}$, for $R^{\text{deg-3}}_{q,q}(x)$, $q \in \{7, 8, 9, 11, 13\}$, and for $R^{\text{1-pt}}_{8,2}(x)$, $R^{\text{1-pt}}_{16,2}(x)$, $R^{\text{deg-3}}_{4,2}(x)$, $R^{\text{deg-3}}_{8,2}(x)$, and $R^{\text{deg-3}}_{16,2}(x)$.*

2. *For the missing cases $R_{2,2}^{\textit{1-pt}}(x)$, $R_{3,3}^{\textit{1-pt}}(x)$, $R_{2,2}^{\textit{deg-3}}(x)$, $R_{3,3}^{\textit{deg-3}}(x)$, $R_{4,4}^{\textit{deg-3}}(x)$, and $R_{5,5}^{\textit{deg-3}}(x)$, the best fitting distribution was the gamma distribution.*

3. *The second best fitting distribution was the extreme value distribution for $R_{3,3}^{\textit{1-pt}}(x)$, $R_{3,3}^{\textit{deg-3}}(x)$, $R_{4,4}^{\textit{deg-3}}(x)$, $R_{5,5}^{\textit{deg-3}}(x)$.*

Our results show that for $q \geq 3$, among the two-parameter distributions, the extreme value distribution function is a reasonable estimation of the rate function of subfield subcodes of Hermitian codes.

The extreme value distribution is also referred to as Gumbel or type 1 Fisher-Tippet distribution. In probability theory, these are the limiting distributions of the minimum of a large number of unbounded identically distributed random variables. The extreme value distribution function is

$$F(x; \alpha, \beta) = 1 - \exp\left(-\exp\left(\frac{x-\alpha}{\beta}\right)\right),$$

with location parameter $\alpha \in \mathbb{R}$ and a scale parameter $\beta > 0$. In figures 3 and 4, we visualized the fitting of the extreme value distribution function to our experimental results on the true dimension of subfield subcodes of Hermitian codes.

The occurrence of the extreme value distribution in the context of subfield subcodes of AG codes may be somewhat surprising, and we cannot give an understandable mathematical explanation for this. However, the rank of random matrices over finite fields is known to be related to the class of Gumbel type distributions; see Cooper's result [Coo00, Theorem 2] for the theoretical background. This theory has been applied to parameter estimates of random erasure codes by Studholme and Blake [SB10].

5 McEliece cryptosystem: attacks and applications

The last chapter provides the first step of our future work toward security analysis of McEliece cryptosystem based on Hermitian subfield subcodes. In the long term, we aim to make a comprehensive study in which we measure the McEliece cryptosystem security. Our attempt intends to improve the practicality of the underlying cryptosystem.

To assess the security of McEliece cryptosystem, we present some well-known attacks, for the reason that one of the security measurements of a cryptographic scheme is its resistance to standard cryptanalysis. The structure of this chapter is the following: we start with an overview of post-quantum cryptography [Nis; Aru+19]. We present an application of the subfield subcodes of Hermitian codes to cryptography. Mainly, we give a formula of the public key size in terms of the code rate using the result of section 4.

5.1 Post-quantum cryptography

In 1994 Shor [Sho94] introduced a quantum algorithm that is efficient in breaking cryptosystems which are believed to be secure for classical computers. Recently, the most

Figure 3: Estimating the extended rate function by extreme value distribution for subfield subcodes of 1-point Hermitian codes

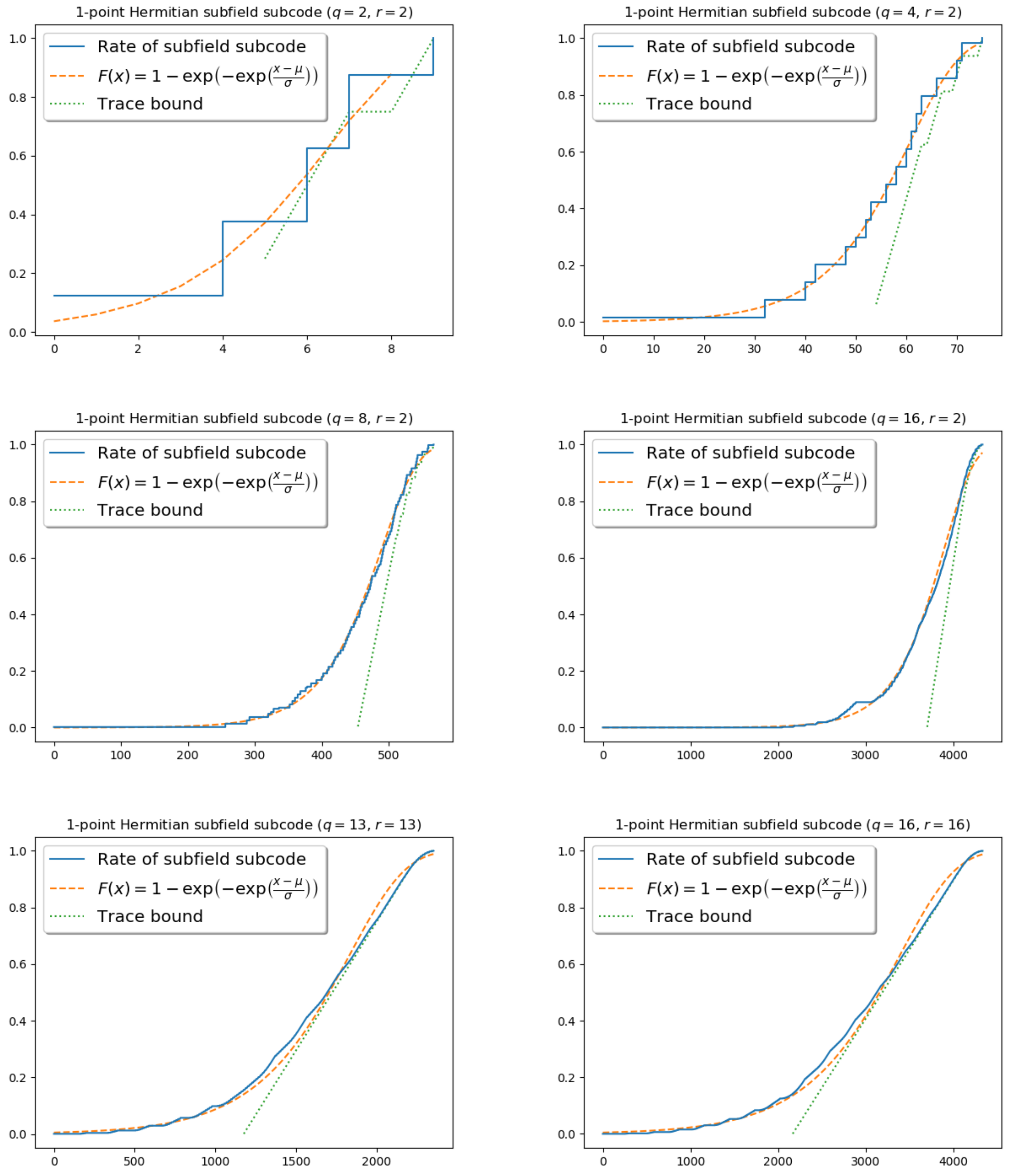
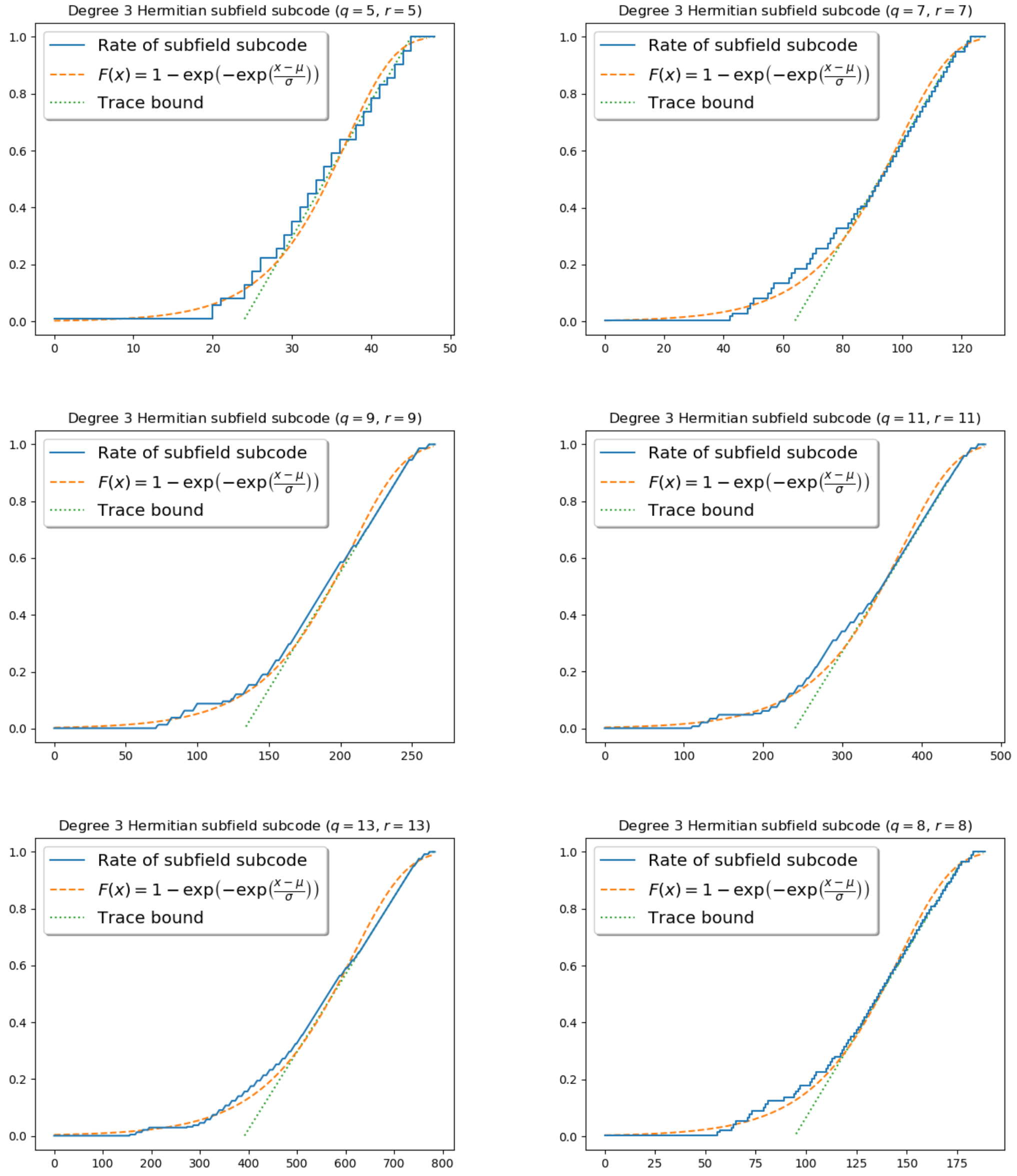


Figure 4: Estimating the extended rate function by extreme value distribution for subfield subcodes of degree 3 Hermitian codes by extreme value distribution



frequent question is what sort of cryptosystems we can use in the presence of quantum computers. Once these latest will be available, we must have systems that are part of post-quantum techniques and which are known as post-quantum cryptography. It consists of different classes.

Replacing these alternative systems will take time. Moreover, quantum-resistant cryptosystems should be in today's use to protect sensitive data. The construction of a secure cryptographic scheme must rely on a computationally hard problem. In classical cryptography, there are many schemes in which security is based on a difficult problem that a classical computer cannot solve, but a quantum computer can.

5.2 Code-based cryptography

Code-based cryptography is a set of cryptosystems in which the underlying trapdoor function is based on error-correcting codes. The first code-based cryptosystem was introduced by Robert J. McEliece in 1978. One must randomly select an error-correcting code to generate the private key that is the structure of the chosen code and the public key whose generator matrix has been randomly permuted. The plaintext is a codeword to which we add some errors in order to get ciphertext. Only the private key's possessor can decode the ciphertext to remove errors and recover the plaintext. It is required to adjust some parameters that concern its efficiency. Until now, there is no serious attack that threatens the security of the McEliece scheme even on quantum computers.

5.3 Attacks against code-based cryptography

In the literature, several attacks have been proposed against McEliece cryptosystem in general, and against McEliece systems that are based on AG codes in specific, see [BBC13]. Attacks can be divided into two classes: *structural* or key recovery attacks which aimed at recovering the secret code, and *decoding*, or message recovery attacks that seek to decrypt the transmitted ciphertext. The generic decoding attack against the McEliece scheme is the information set decoding (ISD) algorithm. The most recent and most effective structural attack against AG code-based McEliece systems is the Schur product distinguisher, which is given in [CMCP17], where the authors show that subfield subcodes of AG codes still resist. We focus on attacks based on Information Set Decoding (ISD) since they are useful for our case, and also, it is assumed to have the lowest complexity [Nie+12].

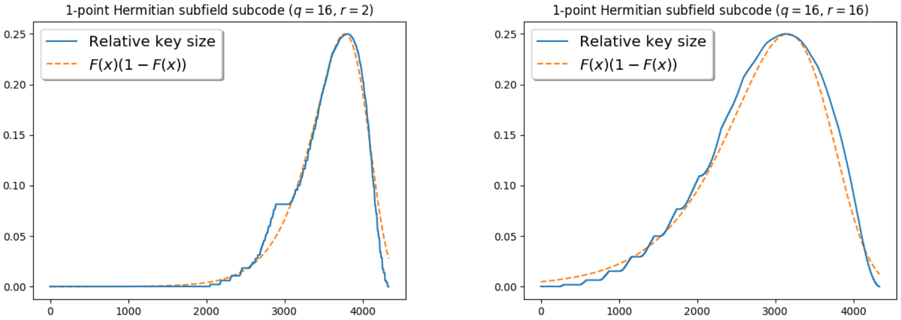
5.4 Selecting parameters to secure McEliece cryptosystem

We apply the result concerning the estimation of the true dimension of the subfield subcodes of Hermitian codes to estimate the key size of the McEliece cryptosystem. The largest (but not the only) part of the public key of the McEliece cryptosystem is the matrix G. G is either the $n \times k$ generator matrix, or the $n \times (n-k)$ parity check matrix. In either case, G may be assumed to be in a systematic form, which means that the public key is given by $k(n-k)$ elements of $\mathbb{F}_r$. Hence, the key size is

$$\log_2(r)k(n-k).$$

In particular, for a fixed field $\mathbb{F}_r$ and length n, the key size is proportional to $R(1-R)$, see [Nie+12]. The true values of $R^{\gamma}_{q,r}(s)(1-R^{\gamma}_{q,r}(s))$ can be estimated by $F(x)(1-F(x))$, where $F(x)$ is the extreme value distribution function [EKN20], see Figure 5.

Figure 5: Estimating the key size $n^2R(1-R)$



References

[Aru+19] F. Arute et al. "Quantum supremacy using a programmable superconducting processor". In: *Nature* 574.7779 (2019), pp. 505–510.

[BBC13] M. Baldi, M. Bianchi, and F. Chiaraluce. "Security and complexity of the McEliece cryptosystem based on quasi-cyclic low-density parity-check codes". In: *IET Information Security* 7.3 (2013), pp. 212–220.

[BS95] S. V. Bezzateev and N. A. Shekhunova. "Subclass of binary Goppa codes with minimal distance equal to the design distance". In: *IEEE Trans. Inform. Theory* 41.2 (1995), pp. 554–555.

[Cas20] F. de Castro. *fitmethis, Version 1.3.0.0*. MATLAB Central File Exchange. Jan. 2020.

[Coo00] C. Cooper. "On the distribution of rank of a random matrix over a finite field". In: *Proceedings of the Ninth International Conference "Random Structures and Algorithms" (Poznan, 1999)*. Vol. 17. 3-4. 2000, pp. 197–212.

[CKT99] A. Cossidente, G. Korchmáros, and F. Torres. "On curves covered by the Hermitian curve". In: *J. Algebra* 216.1 (1999), pp. 56–76.

[CMCP17] A. Couvreur, I. Márquez-Corbella, and R. Pellikaan. "Cryptanalysis of McEliece cryptosystem based on algebraic geometry codes and their subcodes". In: *IEEE Trans. Inform. Theory* 63.8 (2017), pp. 5404–5418.

[Del75] P. Delsarte. "On subfield subcodes of modified Reed-Solomon codes". In: *IEEE Trans. Information Theory* IT-21.5 (1975), pp. 575–576.

[EKN19] S. El Khalfaoui and G. P. Nagy. “On the dimension of the subfield subcodes of 1-point Hermitian codes”. In: *Advances in Mathematics of Communications* 0.0 (2019), p. 0. arXiv: `arxiv:1906.10444 [math.AG]`.

[EKN20] S. El Khalfaoui and G. P. Nagy. “Estimating The Dimension Of The Subfield Subcodes of Hermitian Codes”. In: *Acta Cybernetica* (2020). To appear. arXiv: `arxiv:2004.05896 [math.AG]`.

[Gap] *GAP – Groups, Algorithms, and Programming, Version 4.10.2.* The GAP Group, June 2019.

[KN13] G. Korchmáros and G. P. Nagy. “Hermitian codes from higher degree places”. In: *J. Pure Appl. Algebra* 217.12 (2013), pp. 2371–2381.

[LC01] S. Lin and D. J. Costello. *Error control coding.* Vol. 2. Prentice hall, 2001.

[Men+13] A. J. Menezes et al. *Applications of finite fields.* Vol. 199. Springer Science & Business Media, 2013.

[NEK19] G. P. Nagy and S. El Khalfaoui. *HERmitian, Computing with divisors, Riemann-Roch spaces and AG-odes of Hermitian curves, Version 0.1.* GAP package. Mar. 2019.

[Nie+12] R. Niebuhr et al. “Selecting parameters for secure McEliece-based cryptosystems”. In: *International Journal of Information Security* 11.3 (June 2012), pp. 137–147.

[PJ14] F. Piñero and H. Janwa. “On the subfield subcodes of Hermitian codes”. In: *Designs, codes and cryptography* 70.1-2 (2014), pp. 157–173.

[Nis] *Post-Quantum Cryptography.* `http://csrc.nist.gov/projects/post-quantum-cryptography`. Updated: March 25, 2020.

[Ros+92] A. M. Roseiro et al. “The trace operator and redundancy of Goppa codes”. In: *IEEE Trans. Inform. Theory* 38.3 (1992), pp. 1130–1133.

[SMS97] T. Shibuya, R. Matsumoto, and K. Sakaniwa. “An improved bound for the dimension of subfield subcodes”. In: *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences* 80.5 (1997), pp. 876–880.

[Sho94] P. Shor. “Polynomial-time algorithm for prime factorization and discrete logarithms on a quantum computer: proc”. In: *35th Annual Symp. on the Foundations of Computer Science.* Vol. 124. 1994.

[Ste12] S. A. Stepanov. *Codes on algebraic curves.* Springer Science & Business Media, 2012.

[Sti90] H. Stichtenoth. “On the dimension of subfield subcodes”. In: *IEEE Transactions on Information Theory* 36.1 (1990), pp. 90–93.

[Sti09] H. Stichtenoth. *Algebraic function fields and codes.* Vol. 254. Springer Science & Business Media, 2009.

[SB10] C. Studholme and I. F. Blake. “Random matrices and codes for the erasure channel”. In: *Algorithmica* 56.4 (2010), pp. 605–620.

[TM19] I. The MathWorks. *Statistics and Machine Learning Toolbox.* Natick, Massachusetts, United State, 2019.

[VDV91] M. Van Der Vlugt. “A new upper bound for the dimension of trace codes”. In: *Bulletin of the London Mathematical Society* 23.4 (1991), pp. 395–400.

[Vér98] P. Véron. “Goppa codes and trace operator”. In: *IEEE Trans. Inform. Theory* 44.1 (1998), pp. 290–294.

[Véro01] P. Véron. “True dimension of some binary quadratic trace Goppa codes”. In: *Des. Codes Cryptogr.* 24.1 (2001), pp. 81–97.

[Vér05] P. Véron. “Proof of conjectures on the true dimension of some binary Goppa codes”. In: *Des. Codes Cryptogr.* 36.3 (2005), pp. 317–325.

[Vlu91] M. van der Vlugt. “On the dimension of trace codes”. In: *IEEE Transactions on Information Theory* 37.1 (1991), pp. 196–199.