UNIVERSITY OF UTRECHT

MASTER THESIS

Skew Coding and Skew Factorisation

Author: Nela Lekić
Student number: 3022420
Supervisor: Prof. dr. Frits Beukers
Co-supervisor: Dr. Wilberd van der Kallen

July 7, 2011

Abstract

The thesis consists of two parts. First part is on coding theory, more specifically on skew codes. It starts with a short introduction to coding theory by presenting a few codes and giving bounds on their quality. It also includes an overview of papers by Boucher and Ulmer on cyclic codes over skew polynomial rings that motivated the thesis. The second part deals with factorisation in skew polynomial rings. We improve the estimate of the bound of a polynomial in skew rings given by Boucher and Ulmer and present a new approach to factorisation using difference operators.

Contents

1 Introduction **4**
- 1.1 Block Codes . . . 5
- 1.2 The Original Hamming Code . . . 7
- 1.3 Linear Codes . . . 8
- 1.4 Cyclic Codes . . . 9
- 1.5 BCH Codes . . . 12
- 1.6 Bounds on Codes . . . 13

2 Skew Cyclic Codes **16**
- 2.1 Skew Polynomial Rings . . . 16
- 2.2 Characterization of two-sided ideals . . . 18
- 2.3 Codes over Skew Polynomial Rings . . . 19
- 2.4 Skew Codes . . . 20
- 2.5 Bounds on Polynomials . . . 21

3 Factorisation in Commutative Polynomial Rings **24**
- 3.1 Berlekamp's algorithm . . . 24
- 3.2 Factorisation of $x^n - 1$ . . . 26

4 Factorisation in Skew Polynomial Rings **28**
- 4.1 Difference Operators . . . 28
- 4.2 Cayley-Hamilton and Implications . . . 32
- 4.3 Factorisation of $x^n - 1$ . . . 35
- 4.4 Jordan-Hölder and Factorisation in the Center . . . 39

Chapter 1

Introduction

The problem of reliable communication is a very old one. Coding theory on the other hand is rather young. It was born in a now classic paper from 1948 "A mathematical theory of communication" by Shannon. The model is as follows. A sender wants to communicate a message to the receiver. Rather than sending it directly, the sender encodes the message and sends it through the noisy communication channel. When the message is received it might contain errors. We would like to prevent errors like this occurring in the digital world and somehow make sure we received the right message before we decode it. This groundbreaking idea of error correcting codes is due to Hamming. He developed it in his famous paper "Error detecting and error correcting codes" in 1950. The coding theory is thus concerned with developing codes that have efficient encoding and decoding algorithms as well as the ability to detect and correct the errors in the communication.

The goal of the thesis was initially to study so called skew codes, the codes over non-commutative polynomial rings which are a generalization of the usual ring of polynomials. This was motivated by a paper of Boucher, Geiselmann and Ulmer [2] where they introduce them. The idea was to construct the codes and find the ones with good properties. This is (partially) dealt with in chapters 1 and 2. At the end of the chapter 2 we encountered an interesting question which shifted the goal of the thesis. The question arose from a practical problem of constructing a large number of codes in an efficient way, and to answer it we had to look into factorisation of polynomials in non-commutative rings. That became the topic of the second part of the thesis.

Chapter 1 serves as a quick introduction to basic principles of coding theory. It introduces a large family of linear codes, and within them, the cyclic codes and BCH codes. The study of skew codes starts in chapter 2. There we outline the properties of skew rings, define codes over them, and finally discuss bounds on polynomials, which give motivation for the next two chapters. Chapter 3 is an overview of already well known facts about factorisation of polynomials in commutative polynomial rings. The purpose of this chapter was to gain insight into factorisation in non-commutative rings studied in chapter 4. There we develop an approach that uses difference operators. The central idea of this approach is to use the one-to-one correspondence between right factors of a polynomial and vector subspaces of its solution space. In that way we were

able to answer the question posed in chapter 2.

1.1 Block Codes

Van Lint gives a nice introduction to coding theory in [12]. Most of the definitions and lemmas in this chapter are taken from that book.

By a *code* we will always mean a block code. A block code is a set of words (codewords, blocks) of length n, that take entries from an alphabet Σ with q symbols and can be decoded independently from each other. If we denote by Σ^n a set of all possible words of length n that take entries from Σ, then a *block code* C is a subset of Σ^n.

Codes that are not block codes, i.e. that have words which are not of constant length, are called *convolutional* codes. For an introduction to convolutional codes see [13] and for a comparison with block codes see [13, Section 11.4].

We start by listing the most important definitions that we will use throughout the text.

Definition 1. The Hamming distance $d(x, y)$ between two words $x, y \in \Sigma^n$ is defined as a number of positions in which they differ:

$$d(x, y) := \#\{i : x_i \neq y_i, 1 \leqslant i \leqslant n\}$$

It can easily be checked that Hamming distance is a metric on Σ^n:

(i) $d(x, y) = d(y, x)$

(ii) $d(x, z) \leqslant d(x, y) + d(y, z)$

(iii) $d(x, y) = 0 \Leftrightarrow x = y$

We will especially be interested in the minimum Hamming distance between pairs of codewords in a code.

Definition 2. The minimum distance of code C is

$$\min\{d(x, y) : x, y \in C, x \neq y\}.$$

Another important concept is the *Hamming weight*.

Definition 3. The Hamming weight $w(x)$ of a codeword x is the number of non-zero coordinates of x.

Definition 4. A code that is able to detect up to e errors is called *e-error-detecting*. If it can correct up to e errors, we call such a code *e-error-correcting*.

A code with a minimum distance $d = 2e + 1$ is $2e$-error detecting, since if we were to receive a word with $2e$ errors it would simply not be a codeword. A code that is able to correct up to e errors must have a minimum distance of at least $d = 2e + 1$, because if any two codewords differ in at least $2e + 1$ positions, the received word with e errors

or less resembles the intended codeword more then it resembles any other word. If an e-error-correcting code has a property that every $\mathbf{x} \in \Sigma^n$ has distance $\leqslant e$ to exactly one codeword, then we call such a code *perfect*. For perfect e-error correcting codes we have the following immediate result.

Theorem 1 (Sphere-packing condition)**.** *Let C be a perfect e-error correcting code over Σ^n, with Σ an alphabet with q symbols. Then*

$$|C| \sum_{i=0}^{e} \binom{n}{i} (q-1)^i = q^n.$$

The intuition behind this is that we can think of Σ^n as a large box into which we want to pack spheres. A sphere of radius e that is centered at c is a set of words x in Σ^n that have $d(x, c) \leqslant e$. In order to have an e-error correcting code we can choose as codewords the centers of spheres that have a radius e. If we want a code to be perfect, then those spheres need to be pairwise disjoint. The sphere packing condition then says that for perfect codes we can pack the spheres centered at codewords in such a way that they completely fill the volume of the box Σ^n.

When a code C has many codewords it is able to carry a large number of possible messages. But the more words it has the harder it becomes to distinguish between them if an error in transmission occurs. On the other hand, if a code has very few codewords, it carries little information, but it is easier to correct errors. It becomes natural then to define the *information rate* as follows.

Definition 5. Information rate of a code C of length n over an alphabet with q symbols is $\dfrac{\log_q |C|}{n}$.

Example 1. *Consider two extreme cases. A code C_1 that consists of a single word carries no information but we can never confuse one word for another since there only exists one. Its information rate is, as expected, equal to zero.*

$$\frac{\log_q |C_1|}{n} = \frac{\log_q 0}{n} = 0$$

A code C_2, on the other hand, that as its words has all possible words of length n given an alphabet with q symbols has information rate

$$\frac{\log_q |C_2|}{n} = \frac{\log_q q^n}{n} = 1$$

but a single error leads to another existing codeword, so we can never hope to correct it.

The last thing we will mention in this section are the assumptions that we made on the communication channel. The decoding principle that is used is called *maximum-likelihood principle* and it says that if we receive a word x we want to find a codeword c such that $d(x, c)$ is minimized. This is a consequence of two assumptions: the first one is that all codewords are equally likely, and the second is that an error pattern with t_1 errors is more likely then the one with t_2 errors if $t_1 < t_2$.

1.2 The Original Hamming Code

Hamming had an access to an early electronic computer. He wrote his code on paper-tape. Paper tape had eight places per line and on each place there could be a hole punched or not. Seven of those eight places carried information, the last bit was a *parity check*. That is to say, it was an extra bit that carried no information but was added to ensure that the number of bits with the value one in each line is even. For example if the number of information bits with value one was odd, then the parity bit would have value one, so that the overall parity is even. In that way a machine could detect at most one error (if for example a line with odd number of bits of value one is received), but could not fix it. Hamming developed a code that would be able to do precisely that in his famous paper [7].

His original code is a binary code of length 7. The codewords are vectors $\mathbf{c} = (c_1, ..., c_7)$ in $\mathbb{F}_2^7$ that satisfy the following set of equations:

$$\begin{aligned} c_1 + c_3 + c_5 + c_7 &= 0 \\ c_2 + c_3 + c_6 + c_7 &= 0 \\ c_4 + c_5 + c_6 + c_7 &= 0 \end{aligned}$$

To encode a sequence of bits, we cut it into 4 bit blocks because from the system of equations above we see that if we fix c_3, c_5, c_6 and c_7, then c_1, c_2 and c_4 are completely determined. The information rate is thus $\frac{4}{7}$. It is not hard to see that the minimum distance is 3 and thus Hamming code is 1-error correcting. To see this define the *syndrome* $(z_1, z_2, z_3) \in \mathbb{F}_2^3$ to be

$$\begin{aligned} z_1 &= x_1 + x_3 + x_5 + x_7 \\ z_2 &= x_2 + x_3 + x_6 + x_7 \\ z_3 &= x_4 + x_5 + x_6 + x_7. \end{aligned}$$

If $x \in \mathbb{F}_2^7$ is a codeword, then $(z_1, z_2, z_3) = 0$. But if instead of a codeword we receive a string $x \in \mathbb{F}_2^7$ such that it differs from some codeword c in only one position, then we can tell from the syndrome the exact position in which an error has occurred. Namely, if (z_1, z_2, z_3) is the syndrome of x then the error is in position $z_1 + 2z_2 + 4z_4$.

Hamming codes can be generalized to describe a whole family of certain linear codes over $\mathbb{F}_q$ that are all 1-error correcting. To define the general Hamming codes we will need a concept of parity check matrix. This naturally arises in the context of linear codes and is explained in the next section. However, the following theorem is a nice application of the sphere-packing condition that I wanted to include here. So for this purpose a parity check matrix is simply a matrix representation of the system of linear equations that define the codewords. For example, for the original Hamming code we have

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

A string $x \in \mathbb{F}_2^7$ is a codeword if and only if

$$xH^T = 0.$$

Definition 6. Let k be a strictly positive integer and $n = \frac{q^k-1}{q-1}$. Consider the column vector space $\mathbb{F}_q^k$. Let H denote a $k \times n$ matrix whose columns are the $\frac{q^k-1}{q-1}$ distinct (up to scalars) non-zero vectors of $\mathbb{F}_q^k$. Then the Hamming $[n, n-k]$ code is a linear code of length n with H^T as a parity check matrix.

Theorem 2 (cf. Theorem (3.3.2) in [12])**.** *Hamming codes are perfect.*

Proof. By definition of a Hamming code of length n and dimension k over $\mathbb{F}_q$ we have $n = \frac{q^k-1}{q-1}$. We know Hamming codes are 1-error correcting, so we want to consider disjoint spheres of radius one centered around codewords of C. Let c be a codeword. Then the number of n-tuples in $\mathbb{F}_q$ in a sphere of radius one centered around c, $|B_1(c)|$, is

$$|B_1(c)| = \binom{n}{0} + \binom{n}{1} = 1 + n(q-1) = q^k.$$

The number of codewords of C is q^{n-k}. Then $q^{n-k}q^k = q^n$, which is all of the $\mathbb{F}_q^n$. Thus C is perfect. □

1.3 Linear Codes

To introduce some algebraic structure into codes, we let the alphabet Σ be a finite field $\mathbb{F}_q$ with q a prime power. The space Σ^n then becomes an n-dimensional vector space $\mathbb{F}_q^n$ over $\mathbb{F}_q$.

Definition 7. Let $\mathbb{F}_q$ be a field with q elements and n an integer. A linear code C of length n and dimension k is a k dimensional linear subspace of $\mathbb{F}_q^n$. We use notation $[n, k]$ to refer to such a code.

In other words, a linear code C is a subset of $\mathbb{F}_q^n$ such that

(i) $\mathbf{0} \in C$

(ii) if $\mathbf{x}, \mathbf{y} \in C$, then $\mathbf{x} + \mathbf{y} \in C$

(iii) for all $\mathbf{x} \in C$ and $\lambda \in \mathbb{F}_q$ we have $\lambda\mathbf{x} \in C$

Codewords of a linear code are thus n-tuples over $\mathbb{F}_q$. An advantage of a linear code over a nonlinear one is that is it very easy to represent it. For a linear $[n, k]$ code C any $k \times n$ matrix whose rows form a basis for C completely determines the code. We call such a matrix a *generating matrix* of a code C.

We saw that the information rate of a code was defined as $\frac{\log_q |C|}{n}$. For a linear code the number of codewords is $|C| = q^k$ so the information rate then simplifies to $\frac{k}{n}$. Another simplification we gain from imposing structure on a code is in computing its minimum distance. For an arbitrary, unstructured code one must check the distances between all possible pairs of words in order to find the minimum distance. That is, one must compute $\binom{|C|}{2}$ distances.

Theorem 3. *Let C be a linear code. Then its minimum distance equals its minimum weight.*

Proof. Note that the weight function of a codeword $\mathbf{x}$ was defined as the number of non-zero coordinates of $\mathbf{x}$ and so we have $w(\mathbf{x}) = d(\mathbf{0}, \mathbf{x})$. Then

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= d(\mathbf{x} - \mathbf{y}, \mathbf{0}) \\ &= w(\mathbf{x} - \mathbf{y}) \\ &= w(\mathbf{z}) \end{aligned}$$

Since $\mathbf{x}$ and $\mathbf{y}$ are in C which is linear, it follows that $\mathbf{z}$ is also a word in C. □

Once we know the minimum distance d, notation $[n, k, d]$ is a common way to refer to a linear code of length n, dimension k and minimum distance d.

Definition 8. Let C be a $[n, k]$ code. Its dual code $C^{\perp}$ is defined as

$$C^{\perp} := \{y \in \mathbb{F}_q^n : \langle x, y \rangle = 0 \ \forall x \in C\}.$$

The dual code $C^{\perp}$ is a linear $[n, n-k]$ code. For every $y \in C^{\perp}$ the equation $\langle x, y \rangle = 0$ holds for every $x \in C$ by definition, and we call it *parity check equation*. The generator matrix H of $C^{\perp}$ is called a *parity check matrix* for C. Denote by G the generator matrix of C. When G is in reduced echelon form then

$$G = (I_k \ P)$$

since G is a $k \times n$ matrix. In order to find the generator matrix H for $C^{\perp}$ note that we want H to be a $(n-k) \times n$ matrix such that $GH^T = 0$. Thus

$$H = (-P^T \ I_{n-k}).$$

Note also that from $GH^T = 0$ we get that $x \in \mathbb{F}_q^n$ is a codeword of C if and only if $xH^T = 0$. So we get $n-k$ linear equations that every codeword of C needs to satisfy.

1.4 Cyclic Codes

Cyclic codes are a small subset of the set of linear codes. They are the most common block codes used in practice. There are a few reasons cyclic codes are nice to study. One is that they have a very rich algebraic structure, and another is that many important codes (BCH codes for example) are cyclic.

Definition 9. A code C of length n is called cyclic if for every codeword

$$c = (c_0, c_1, ..., c_{n-1}) \text{ in } C,$$

we have that

$$c' = (c_{n-1}, c_0, ..., c_{n-2}) \text{ is also in } C.$$

Even though we said cyclic codes were linear, it is not clear why that is the case from this definition. In principle it is possible to have nonlinear cyclic codes since the way we defined them does not require linearity. However, because of the advantages of imposing linearity it is common to only consider linear cyclic codes.

For a given $c \in C$ any number of right or left shifts on c by definition gives another word in C. Any linear combination of shifted versions of c also gives a codeword. This suggest the following construction. Let G denote a set of all possible right shifts of a word c. Then the linear span of G is the smallest linear cyclic code C containing c. By this construction it is clear that a single word determines a code. We call such a word a *generator*. A generator need not be unique. This deserves a more precise treatment. In order to do that it is convenient to think of codewords as polynomials in the following way. Let $\mathbb{F}_q$ denote a finite field of q elements and $\mathbb{F}_q[x]$ a ring of polynomials in x with coefficients in $\mathbb{F}_q$. To every codeword

$$c = (c_0, c_1, ..., c_{n-1}) \in \mathbb{F}_q^n$$

we associate the *code polynomial*

$$c(x) = c_0 + c_1 x + ... + c_{n-1} x^{n-1} \in \mathbb{F}_q[x]/(x^n - 1).$$

With this convention established we can sometimes abuse the notation and call a code polynomial $c(x)$ a codeword in C.

Note that a shifted codeword $c' = (c_{n-1}, c_0, ..., c_{n-2})$ in C has associated polynomial

$$c'(x) = c_{n-1} + c_0 x + ... + c_{n-2} x^{n-1}$$

and that $c'(x) = xc(x)$ modulo $x^n - 1$. In this way we can represent a single right cyclic shift between two codewords in C with a multiplication by x in a ring of polynomials modulo $x^n - 1$.

We know that applying any number of cyclic shifts on $c \in C$ gives us another codeword

$$x^i c(x) \mod x^n - 1 \in C$$

and that any linear combination of words in C produces another word in C

$$\sum_{i=0}^{d} a_i x^i c(x) \mod x^n - 1 \in C$$

where $a_i \in \mathbb{F}_q$. In other words, for any polynomial $a(x) \in \mathbb{F}_q[x]/(x^n - 1)$ and any codeword $c(x) \in C$ the product $a(x)c(x)$ is also in C. This proves the only if direction of the following:

Theorem 4. *A linear code C in $\mathbb{F}_q^n$ is cyclic if and only if C is an ideal in $\mathbb{F}_q[x]/(x^n - 1)$.*

Proof. What's left to show is that if C is an ideal in $\mathbb{F}_q[x]/(x^n - 1)$ then C is cyclic. Let c be any codeword. Then a single right shift, represented by $xc(x) \mod x^n - 1$, is a codeword of C and thus C is cyclic. □

We saw that a cyclic code is generated by a single codeword. We choose this word to be a monic of smallest degree and call it the *generator polynomial*.

Proposition 1. *Let $g(x)$ be a generator polynomial of a cyclic code. Then $g(x)$ divides $x^n - 1$.*

Proof. Suppose $g(x)$ does not divide $x^n - 1$. By the division algorithm we have that there exist $f(x), r(x) \in \mathbb{F}_q[x]$ such that

$$x^n - 1 = f(x)g(x) + r(x),$$

with $deg(r(x)) < deg(g(x))$. Then

$$\begin{aligned} r(x) &= x^n - 1 - f(x)g(x) \\ &\equiv -f(x)g(x) \mod x^n - 1 \end{aligned}$$

is also in C but contradicts the minimality of $g(x)$. It must therefore be that $r(x) = 0$ and $g(x)$ divides $x^n - 1$. □

This gives us an easy way to construct cyclic codes. Let $x^n - 1 = f_1(x)...f_m(x)$ be a decomposition over $\mathbb{F}_q$ into irreducible factors f_i. Then we know that each f_i, but also the product of each combination of f_i's, is a generator polynomial of some cyclic code of length n.

Furthermore, if a polynomial $g(x)$ of degree r is a generator of a cyclic code of length n then the basis for C is given by the following set:

$$g(x), xg(x), x^2g(x), ..., x^{n-r-1}g(x).$$

Let $g(x) = g_0 + g_1x + ... + g_rx^r$. The generating matrix for this code is

$$G = \begin{bmatrix} g_0 & g_1 & ... & g_r & 0 & 0 & ... & 0 \\ 0 & g_0 & ... & g_{r-1} & g_r & 0 & ... & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & ... & g_0 & g_1 & g_2 & ... & g_r \end{bmatrix}.$$

This is a $(n-r) \times n$ matrix and thus a polynomial $g(x)$ of degree r generates a $[n, n-r]$ cyclic code.

By proposition 1, a generator $g(x)$ of a cyclic code C divides $x^n - 1$. Then there exists a polynomial $h(x)$ such that $x^n - 1 = g(x)h(x)$. Note that in ring $\mathbb{F}_q[x]/(x^n - 1)$ we have $g(x)h(x) = 0$. For every codeword $c(x) \in C$ by theorem 4 we have $c(x)h(x) = c'(x)g(x)h(x) = 0$. In other words, $c(x) \in C$ if and only if $c(x)h(x) = 0$. Thus $h(x)$ is a check polynomial for C and the parity check matrix for C is simply

$$H = \begin{bmatrix} 0 & 0 & ... & 0 & h_k & ... & h_1 & h_0 \\ 0 & 0 & ... & h_k & ... & h_1 & h_0 & 0 \\ \vdots & & & & & & & \vdots \\ h_k & & ... & h_1 & h_0 & 0 & ... & 0 \end{bmatrix}.$$

Next we review the class of cyclic codes that is still being used a lot in practice.

1.5 BCH Codes

This class of cyclic codes was defined by Hocquenghem in 1959 and independently by Bose and Ray-Chaudhuri in 1960, thus the name. Practically they are interesting because of a simple decoding procedure that requires only a very simple decoding device rather than a computer. Mathematically they are interesting for their flexibility: apart from sharing many good properties with cyclic codes they allow for a certain control of minimum distance as we shall see in this section.

Definition 10. A BCH code of designed distance δ is a cyclic code of length n over $\mathbb{F}_q$ whose generating polynomial $g(x)$ is a least common multiple of the minimal polynomials of $\beta^l, \beta^{l+1}, ..., \beta^{l+\delta-2}$, where β is a primitive nth root of unity and l some integer.

Usually l in the definition above is taken to be $l = 1$. We call such a code a *narrow-sense* BCH code. If β is a primitive element of $\mathbb{F}_{q^m}$ we call such a code a *primitive* BCH code. Note that if β is a primitive element of $\mathbb{F}_{q^m}$ and an nth root of unity, then $n = q^m - 1$. The following theorem explains in what way we "control" the minimum distance.

Theorem 5 (BCH bound)**.** *Let C be a BCH code with designed distance δ. Then the minimum distance of C is greater then or equal to δ.*

Proof. Let C be a BCH code of designed distance δ. By definition, the generating polynomial $g(x)$ is the least common multiple of the minimum polynomials of $\delta - 1$ consecutive powers of a primitive nth root of unity β. Denote them by $\beta^l, \beta^{l+1}, ..., \beta^{l+\delta-2}$ as in the definition. Since C is cyclic, every codeword c of C is a multiple of $g(x)$, and thus $c \in C$ if and only if $c(\beta^i) = 0$ for $i = l, l+1, ..., l+\delta-2$. Suppose all the zeros of the generating polynomials lie in some field $\mathbb{F}_{q^m}$ containing $\mathbb{F}_q$. We can represent $\mathbb{F}_{q^m}$ as a vector space $\mathbb{F}_q^m$ and each β^i as an m-tuple over $\mathbb{F}_q$. The idea is to form a parity check matrix. For each β^i we can form a $m \times n$ matrix whose columns are representations of $1, \beta^i, (\beta^i)^2, ..., (\beta^i)^{n-1}$ in $\mathbb{F}_q^m$. When we put the matrices for each β^i together we get the $m(\delta-1) \times n$ matrix

$$H = \begin{bmatrix} 1 & \beta^l & \beta^{2l} & ... & \beta^{(n-1)l} \\ 1 & \beta^{l+1} & \beta^{2(l+1)} & ... & \beta^{(n-1)(l+1)} \\ \vdots & & & & \vdots \\ 1 & \beta^{l+\delta-2} & \beta^{2(l+\delta-2)} & ... & \beta^{(n-1)(l+\delta-2)} \end{bmatrix}.$$

By construction of H

$$cH^T = 0 \Leftrightarrow c \text{ is a codeword of } C.$$

clearly holds. However, the rows of H are not necessarily independent, and we may delete some to obtain the parity check matrix. We can also consider any $\delta - 1$ columns

of H. The determinant of this $(\delta-1)\times(\delta-1)$ submatrix is a Vandermonde determinant

$$\begin{bmatrix} \beta^{i_1 l} & \dots & \beta^{i_{\delta-1} l} \\ \beta^{i_1 (l+1)} & \dots & \beta^{i_{\delta-1}(l+1)} \\ & & \vdots \\ \beta^{i_1(l+\delta-2)} & \dots & \beta^{i_{\delta-1}(l+\delta-2)} \end{bmatrix} = \beta^{(i_1+\dots+i_{\delta-1})l} \prod_{r>s} (\beta^{i_r} - \beta^{i_s}) \neq 0.$$

We conclude that any $\delta - 1$ columns are linearly independent and thus every codeword has weight at least δ. By theorem 3 from section on linear codes we are done. □

Example 2 (An exercise from [12], chapter 6)**.** *Consider a binary cyclic code C of length $n = 2^m - 1$ where m is odd. Let β be a primitive element of $\mathbb{F}_{2^m}$ and $g(x)$ a generator such that $g(\beta) = g(\beta^{-1}) = 0$. Since $\mathbb{F}_{2^m}$ is a field with characteristic 2 we get $g(\beta^2) = g(\beta)^2 = 0$. For the same reason $g(\beta^{-2}) = 0$ and so $\{\beta^{-2}, \beta^{-1}, \beta, \beta^2\}$ are zeros of the generating polynomial for C. Let C' denote a subcode of C which consists of all the words $c(x)$ that have even weight. For those words $c(\beta^0) = c(1) = 0$ and thus β^0 is a zero of the generating polynomial for C'. Then $\delta \geqslant 6$ and by theorem 5 the minimum distance of C' is at least 6. We conclude that the minimum distance of C is greater than or equal to 5.*

1.6 Bounds on Codes

A natural question that arises after seeing a few different codes is how good they can be. If we set aside efficiency of encoding and decoding algorithms, what are the parameters that tell us when one code is better then the other? We have already seen in section 1.1 that a code with minimum distance d can detect up to $d - 1$ errors and correct up to $\lfloor \frac{d-1}{2} \rfloor$ errors. Obtaining the minimum distance for a general code, however, is computationally expensive. Even linear codes have too many words to just let the computer run through all of them and find one with minimum weight. In this section we will be concerned with finding codes C that given n, q and d have the maximum number of words. To that end, we will discuss upper and lower bounds on $|C|$.

We have already seen this bound in section 1.1. We called it the sphere-packing condition. Implicitly, we used Hamming balls

$$B_r(x) = \{y \in \Sigma^n : d(x, y) \leqslant r\}$$

of radius r centered at x. Denote by $V_q(n, r)$ the number of point in any Hamming ball of radius r over alphabet Σ of size q. Then

$$V_q(n, r) = |B_r(x)| = |B_r(0)| = \sum_{i=1}^{r} \binom{n}{i} (q-1)^i$$

where the first equality holds by definition, second follows from

$$|B_r(x)| = \#\{y : d(x, y) \leqslant r\} = \#\{y - x : d(0, y - x) \leqslant r\} = |B_r(0)|$$

and third is simply a result of counting the points whose distances from 0 are at most r. The **Hamming bound** states

$$|C| \leqslant \frac{q^n}{V_q(n,e)}.$$

For a proof see the "packing argument" of theorem 1.

Next we consider a lower bound on $|C|$. Let C be a largest code among all codes of length n and distance d over alphabet of size q. Denote $|C|$ by $A(n,d)$. The **Gilbert, Shannon, Varshamov bound** is given by

$$A(n,d) \geqslant \frac{q^n}{V_q(n,d-1)}.$$

The proof of this bound essentially states that if we take any word $c \in \Sigma^n$ and keep adding words that are distance d or more from c and all the other words already added, then the code we end up with will have at least $q^n/V_q(n,d-1)$ codewords. To see why this is the case note that the spheres $B_{d-1}(c)$ with $c \in C$ cover Σ^n.

We will now mention two more upper bounds for $A(n,d)$ whose proofs are also elementary. The best known bounds, however, are more complicated and are based on the method by Delsarte which he used to derive the linear programming bound in 1973. For a discussion on the linear programming bound see section 5.3 in [12]. We start with a rather trivial **Singelton bound**.

Theorem 6 (c.f. theorem 5.2.1 in [12])**.** *Let $q,n,d \in \mathbb{N}$, $q \geqslant 2$ and $\theta = 1 - \frac{1}{q}$. Then*

$$A(n,d) \leqslant q^{n-d+1}.$$

Proof. Consider a code C with length n, minimum distance d and $|C| = M$. Remove now the last symbol from every codeword of C. The code we get has the same number of words, length $n-1$ and minimum distance $d-1$. This procedure is called puncturing a code. If we puncture the same code $d-1$ times we get a code of length $n-d+1$ and minimum distance 1 and M words. Now we are done since $M \leqslant q^{n-d+1}$. □

Next theorem is called **Plotkin bound**.

Theorem 7 (c.f. theorem 5.2.3 in [12])**.** *Let $q,n,d \in \mathbb{N}$, $q \geqslant 2$ and $\theta = 1 - \frac{1}{q}$. If $d > \theta n$ then*

$$A(n,d) \leqslant \frac{d}{d-\theta n}.$$

Proof. Let C be a code with length n, minimum distance d and M codewords. Consider all codewords of C as an $M \times n$ matrix. Take any column and let a denote any of q symbols of alphabet Σ. Suppose symbol a occurs m_a times in this column. Then this impacts the sum of distances between all possible $M(M-1)$ ordered pairs of codewords by exactly $m_a(M-m_a)$. If we do this for all symbols of Σ then one column

contributes $\sum_{i=1}^{q} m_i(M - m_i)$ to the sum of all distances. Note that

$$\begin{aligned}
\sum_{i=1}^{q} m_i(M - m_i) &= M^2 - \sum_{i=1}^{q} m_i^2 \ \text{(since } \sum_{i=1}^{q} m_i = M\text{)} \\
&\leqslant M^2 - \frac{1}{q}(\sum_{i=1}^{q} m_i)^2 \ \text{(Cauchy-Schwarz)} \\
&= \theta M^2
\end{aligned}$$

As there are n columns we have $M(M-1)d \leqslant n\theta M^2$, which proves the theorem. $\square$

Chapter 2

Skew Cyclic Codes

So far we looked at cyclic codes that were defined as ideals in quotient ring $\mathbb{F}_q[x]/(x^n - 1)$ with the usual addition and multiplication of polynomials. In this chapter we discuss a generalization of cyclic codes by considering more general polynomial rings with the usual addition of polynomials and non-commutative multiplication. The reason these codes are interesting is that they share most of the properties of cyclic codes and their class is much larger, so the chance of finding good codes is also better. The idea of defining codes over noncommutative polynomial rings was developed in 1985 in a paper of Gabidulin [5]. Boucher, Geiselmann and Ulmer gave a slightly different approach in 2007 in [2]. Later Boucher and Ulmer generalized their approach in [3] to consider an even larger class of codes, not necessarily cyclic, over skew rings.

In this chapter I will start by introducing skew rings and their properties. After that I will look at both skew cyclic and general skew codes as described by Boucher and Ulmer.

2.1 Skew Polynomial Rings

Let $\mathbb{F}_q$ denote a finite field of q elements, θ an automorphism on $\mathbb{F}_q$ and $|\langle\theta\rangle|$ its order. Let $\mathbb{F}_q[x, \theta]$ denote a set of all polynomials with coefficients always written on the left

$$\mathbb{F}_q[x, \theta] = \{a_0 + a_1 x + ... + a_{n-1}x^{n-1} : a_i \in \mathbb{F}_q, n \in N\}.$$

Define $a_0 + a_1 x + ... + a_{n-1}x^{n-1} = b_0 + b_1 x + ... + b_{n-1}x^{n-1}$ if $a_i = b_i \forall i$. Let addition of elements of $\mathbb{F}_q[x, \theta]$ be given by $(a_0 + a_1 x + ...) + (b_0 + b_1 x + ...) = (a_0 + b_0) + (a_1 + b_1)x + ...$ and let multiplication be defined by the rule

$$xa = \theta(a)x.$$

This rule is further extended to all elements of $\mathbb{F}_q[x, \theta]$ by application of the distributive law. Note that multiplication defined in this way is not commutative. The set $\mathbb{F}_q[x, \theta]$ with operations defined as above forms a ring called the *skew polynomial ring* over $\mathbb{F}_q$ with automorphism θ.

Example 3. *Consider $\mathbb{F}_4[x,\theta]$. Then Frobenius automorphism is given by*

$$\begin{aligned}\theta : \mathbb{F}_4 &\to \mathbb{F}_4\\ \alpha &\mapsto \alpha^2.\end{aligned}$$

Let a be a generator of the multiplicative group of $\mathbb{F}_4$. Take $f = x + a$ and $g = ax^2 + 1$ for example. Then

$$\begin{aligned}fg &= (x+a)(ax^2+1)\\ &= xax^2 + x + a^2x^2 + a\\ &= \theta(a)x^3 + a^2x^2 + x + a\\ &= a^2x^3 + a^2x^2 + x + a.\end{aligned}$$

If $f = a_0 + a_1x + ... + a_nx^n$ with $a_n \neq 0$ we say f has degree n. It is not hard to see that for $f, g \in \mathbb{F}_q[x,\theta]$ we have that $deg(fg) = deg(f) + deg(g)$ and $deg(f+g) \leqslant max\{deg(f), deg(g)\}$. This also implies that there are no zero divisors.

Example 4. *Note that $\mathbb{F}_q[x,\theta]$ is not a unique factorisation domain. Consider again $\mathbb{F}_4[x,\theta]$. Listed below are all monic right factors of degree 2 of $x^6 + ax^3$.*

$$\begin{aligned}x^6 + ax^3 &= (x^4 + ax)(x^2)\\ &= (x^4 + ax^3 + x^2)(x^2 + ax)\\ &= (x^4 + ax^3)(x^2 + ax + 1)\end{aligned}$$

Furthermore, $\mathbb{F}_q[x,\theta]$ is a ring endowed with right and left division algorithms. The right division algorithm is analogous to the one in commutative Euclidean domain: given two polynomials $f, g \in \mathbb{F}_q[x,\theta]$ we are looking for $h, r \in \mathbb{F}_q[x,\theta]$ such that

$$f = hg + r \text{ and } deg(r) < deg(g).$$

Polynomials h and r obtained in the right division algorithm are unique in $\mathbb{F}_q[x,\theta]$. Existence of right division implies the existence of right Euclidean algorithm, which in turn implies the existence of *greatest common right divisors* $(gcrd)$ and *least common left multiples* $(lclm)$. The $gcrd$ of f_1 and f_2 is the unique monic polynomial $g \in \mathbb{F}_q[x,\theta]$ of highest degree such that there exist $k_1, k_2 \in \mathbb{F}_q[x,\theta]$ with $f_1 = k_1g$ and $f_2 = k_2g$. The $lclm$ of f_1, f_2 is the unique monic h of lowest degree such that there exist $l_1, l_2 \in \mathbb{F}_q[x,\theta]$ with $h = l_1f_1$ and $h = l_2f_2$.

The left division is similarly defined. Given two polynomials $f, g \in \mathbb{F}_q[x,\theta]$ we are looking for two polynomials $h', r' \in \mathbb{F}_q[x,\theta]$ such that

$$f = gh' + r' \text{ and } deg(r') < deg(g').$$

Definition 11. A left ideal I is a subset of a non-commutative ring R such that I is an additive subgroup of R and for all r in R and all a in I

$$ra \in I.$$

Similarly, a right ideal I is an additive subgroup of R such that for all r in R and all a in I

$$ar \in I.$$

Lemma 1. *Let $\mathbb{F}_q$ be a field with q elements and θ an automorphism. Then every right ideal in $\mathbb{F}_q[x,\theta]$ principal.*

Proof. To see that $\mathbb{F}_q[x,\theta]$ is a principal right ideal domain let I be any of its non-zero right ideals. Let $g \in I$ be a polynomial of least degree not equal to zero. Let f be some polynomial in I. By left division algorithm we have that there exist h and r such that $f = gh + r$ with $deg(r) < deg(g)$. But $r = f - gh$ is in I and so it must be that $r = 0$ by minimality of g in I. Then $f = gh$ and I is a principal right ideal domain. □

Similar argument shows that any left ideal is principal.

2.2 Characterization of two-sided ideals

By lemma 1 all ideals of $\mathbb{F}_q[x,\theta]$ are generated by a single element. We call an ideal I generated by g *two-sided* when it is both a right ideal, $I = g\mathbb{F}_q[x,\theta]$, and a left ideal, $I = \mathbb{F}_q[x,\theta]g$. Understanding which polynomials generate two-sided ideal is important since the generating polynomials of skew codes, as we will see later in this chapter, are exactly the right divisors of polynomials that generate two-sided ideals.

The center $Z(\mathbb{F}_q[x,\theta])$ of $\mathbb{F}_q[x,\theta]$ is the set of all elements that commute with all other elements of $\mathbb{F}_q[x,\theta]$. We call an element $z \in Z(\mathbb{F}_q[x,\theta])$ *central*. Let $\mathcal{F}$ denote the invariant field of θ. Then $\mathcal{F}[x]$ is a commutative subring of $\mathbb{F}_q[x,\theta]$. Let $|\langle\theta\rangle| = m$ and $\mathbb{F}_q[x^m] = \{a_0 + a_1x^m + ... + a_dx^{md} : d \in \mathbb{N}, a_i \in \mathbb{F}_q\}$. A polynomial $f \in \mathbb{F}_q[x,\theta]$ is central if and only if f is both in $\mathcal{F}[x]$ and in $\mathbb{F}_q[x^m]$. In other words, a central element must be of the form $\sum_{i=0}^{d} a_i x^{im}$ where coefficients a_i are in $\mathcal{F}$. Clearly all central elements generate two-sided ideals but there are elements outside the center that also generate two-sided ideals. The following characterizes them.

Lemma 2. *A polynomial $g \in \mathbb{F}_q[x,\theta]$ generates a two-sided ideal if and only if g is of the form $g = x^t h$ with t a fixed integer, $h \in \mathcal{F}[x^m,\theta]$ and m the order of θ.*

Proof. "⇐" First we show that g of such form generates a two-sided ideal. Note that h is a central element and thus (h) is two-sided. It is clear that x^t also generates a two-sided ideal. Then for every $f \in \mathbb{F}_q[x,\theta]$ we have

$$gf = x^t hf = x^t fh = f'x^t h = f'g.$$

for some $f' \in \mathbb{F}_q[x,\theta]$. The first and last equality hold by definition, while the second equality holds because h commutes with all elements of $\mathbb{F}_q[x,\theta]$ and third because (x^t) is two-sided. Similarly, for any s in $\mathbb{F}_q[x,\theta]$ we have

$$sg = sx^t h = x^t s'h = x^t hs' = gs'$$

for some polynomial s' in $\mathbb{F}_q[x,\theta]$. We conclude that (g) is a two sided ideal.

"⇒" Let $g = g'x^t = g_0x^t + g_1x^{t+1} + ... + g_dx^{t+d}$. Since x^t generates a two-sided ideal, it is clear that g generates a two-sided ideal if and only if g' does. Thus we may assume that $g = g' = g_0 + g_1x + ... + g_dx^d$ with $g_0 \neq 0$. So let $g = g_0 + g_1x + ... + g_dx^d$ with $g_0 \neq 0$ be a generator of a two-sided ideal. This

means that for all $a \in \mathbb{F}_q$ there exists $b \in \mathbb{F}_q[x,\theta]$ such that $ag = gb$. In fact, from examining the degrees it follows that $b \in \mathbb{F}_q$. Then from

$$\begin{aligned} ag &= ag_0 + ag_1x + ... + ag_dx^d, \\ gb &= g_0b + g_1xb + ... + g_dx^db \\ &= g_0b + g_1\theta(b)x + ... + g_d\theta^d(b)x^d \end{aligned}$$

we get that $a = b = \theta(b) = \theta^2(b) = ... = \theta^d(b)$. But since a is an arbitrary element of $\mathbb{F}_q$ we must have that all powers of x are multiples of m, the order of θ. Thus $g(x)$ is of the form $g(x) = g_0 + g_1x^m + ... + g_dx^{dm}$.

□

2.3 Codes over Skew Polynomial Rings

We now have all the theory we need to generalize cyclic codes.

Definition 12 (cf. Definition 1 in [2])**.** Let $\mathbb{F}_q$ be a finite field with q elements and θ an automorphism. A linear code C of length n is called θ-cyclic if for every codeword

$$c = (c_0, c_1, ..., c_{n-1}) \in C$$

we have that

$$c' = (\theta(c_{n-1}), \theta(c_0), ..., \theta(c_{n-2})) \in C.$$

Similarly to how polynomial representation of cyclic codes was defined over commutative polynomial rings $\mathbb{F}_q[x]$, here we associate to every word

$$c = (c_0, c_1, ..., c_{n-1}) \in \mathbb{F}_q^n$$

its skew polynomial representation

$$c(x) = c_0 + c_1x + ... + c_{n-1}x^{n-1} \in \mathbb{F}_q[x,\theta]/(x^n - 1).$$

The observation that $c'(x) = c(x)x \mod x^n - 1$ leads to the following important results about the structure of the cyclic code.

Lemma 3 (cf. Lemma 1 in [2])**.** *Let $\mathbb{F}_q$ be a finite field, θ an automorphism and n an integer divisible by the order $|\langle\theta\rangle|$ of θ. Then*

- (i) *The ideal generated by $x^n - 1$ in $\mathbb{F}_q[x,\theta]$ is a two-sided ideal.*
- (ii) *Ring $\mathbb{F}_q[x,\theta]/(x^n - 1)$ is a principal left ideal ring in which ideals are generated by right divisors of $x^n - 1$ in $\mathbb{F}_q[x,\theta]$.*

Proof. Part (i) is immediate. In fact when n divides $|\langle\theta\rangle|$, $x^n - 1$ is a central element. The proof of part (ii) is analogous to proofs we already saw in the commutative case, except that one needs to take care of right and left side. □

The following theorem establishes that θ-cyclic codes, just like cyclic codes in the commutative polynomial rings, are ideals in the residue class ring $\mathbb{F}_q[x,\theta]/$ (x^n-1). The proof is again omitted but is analogous to the one in commutative case.

Theorem 8 (cf. Theorem 1 in [2]). *Let $\mathbb{F}_q$ be a finite field, θ an automorphism and n an integer divisible by the order $|\langle\theta\rangle|$ of θ. A linear code C over $\mathbb{F}_q$ of length n is θ-cyclic if and only if it is a left ideal $(g) \subset \mathbb{F}_q[x,\theta]/(x^n-1)$, generated by a right divisor g of x^n-1.*

Let g be a right divisor of x^n-1 of degree r. Then the θ-cyclic code it generates is a $[n, n-r]$ linear code with generator matrix

$$G = \begin{bmatrix} g_0 & g_1 & \dots & g_r & 0 & \dots & 0 \\ 0 & \theta(g_0) & \dots & \theta(g_{r-1}) & \theta(g_r) & \dots & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & \dots & \theta^{n-r-1}(g_0) & \theta^{n-r-1}(g_1) & \dots & \theta^{n-r-1}(g_r) \end{bmatrix}.$$

Note that in case θ is an identity map we get the class of cyclic codes.

2.4 Skew Codes

Instead of considering the special ideal (x^n-1) we could consider any two-sided ideal of $\mathbb{F}_q[x,\theta]$ and define a θ-code in the following way.

Definition 13 (cf. Definition 1. in [3]). Let $f \in \mathbb{F}_q[x,\theta]$ with $deg(f) = n$ be a generator of a two-sided ideal $I = (f)$ in $\mathbb{F}_q[x,\theta]$. A θ-code of length n is the set of words $c = (c_0, c_1, ..., c_{n-1})$ that are coefficient tuples of elements $c(x) = c_0 + c_1 x + ... + c_{n-1}x^{n-1}$ of a principal left ideal (g) in $\mathbb{F}_q[x,\theta]/I$.

Polynomials that generate two-sided ideals were characterized in section 2.2. From there we know that f in the definition above is of the form $x^t h$ with h a central element. Section 2.3 considered a special f, namely one of the form x^n-1. For the remainder of this chapter we will consider codes in which f is a central element. Such codes are called *θ-central*.

Principal left ideals (g) in $\mathbb{F}_q[x,\theta]/I$, where $I = (f)$ is a two-sided ideal, are generated by right divisors of f. We have seen this argument before: If we let g be a monic of smallest degree and assume that g does not divide f then by division algorithm there exist polynomials h, r such that $f = hg + r$ with $deg(r) < deg(g)$. This is a contradiction on minimality of g, thus $r = 0$, and so g divides f. So to construct a code defined as above we only need one codeword (a right divisor g of f) to specify it. We get all other codewords by taking all left multiples of g in $\mathbb{F}_q[x,\theta]/(f)$.

One way to systematically obtain codes of a given length n is to pick a polynomial f of degree n that generates a two-sided ideal in $\mathbb{F}_q[x,\theta]$. Then and any right divisor g of f is a generator of a θ-code of length n and dimension $n-deg(g)$. The problem here is in finding irreducible right factors of f. Algorithms for factorisation of polynomials in commutative polynomial rings are well known and easy to implement (Berlekamp's algorithm for efficient factorisation of polynomials over finite fields is given in the next

chapter). An algorithm for factorisation in skew polynomial rings does exists (it was introduced by Giesbrecht in [6]), but it is not so straightforward and it is well beyond the scope of this thesis. We will return to the problem of skew factorisation in Chapter 4.

A much quicker but not systematic way to obtain some codes is to pick any element of skew polynomial ring $\mathbb{F}_q[x,\theta]$ and let that element be the generator of a code. In order to construct such a code, we need to know its length. A theorem of Jacobson (Theorem 15 in [8]) states that for every ideal there exists a two-sided ideal contained in it. For our purposes this means that for any $g \in \mathbb{F}_q[x,\theta]$ there exists $f \in \mathcal{F}[x^m]$, with $\mathcal{F}$ the invariant field of θ and m it's order, such that g divides f. The length of a code generated by g is thus $n = deg(f)$. The question that arises is how large can n be. This is an interesting question and it led to development of chapters 3 and 4 and abandoning of computations of codes.

A result of Boucher and Ulmer on maximum length n of a code generated by $g \in \mathbb{F}_q[x,\theta]$ is presented in the next section. In Chapter 4 we improve that result.

2.5 Bounds on Polynomials

Definition 14 (cf. p.38 in [8])**.** Let P be an element of $\mathbb{F}_q[x,\theta]$ and P^* a monic polynomial of minimal degree such that it generates a two-sided ideal (P^*) contained in left ideal (P). The polynomial P^* is called *the bound on* P.

Lemma 4. *Let m be the order of θ, the Frobenius automorphism of $\mathbb{F}_q$. If $P \in \mathbb{F}_q[x,\theta]$ is of degree n, then its bound is of degree $\leqslant m^2 n$.*

Proof. Let $P \in \mathbb{F}_q[x,\theta]$ generate a left ideal. Let P^* be its bound. We construct a polynomial $Q \in \mathbb{F}_q[x,\theta]$ such that $QP = P^*$ generates a two-sided ideal contained in (P). Further, $\mathbb{F}_q$ is an extension field of $\mathbb{F}_p$, say $\mathbb{F}_q = \mathbb{F}_p[\alpha]$, and $[\mathbb{F}_q : \mathbb{F}_p] = m$, so we can write

$$P = \sum_{k=0}^{n} (p_{k,0} + p_{k,1}\alpha + ... + p_{k,m-1}\alpha^{m-1})x^k$$

with $p_{k,i} \in \mathbb{F}_p$ and similarly

$$Q = \sum_{k=0}^{d} (q_{k,0} + q_{k,1}\alpha + ... + q_{k,m-1}\alpha^{m-1})x^k$$

with $q_{k,i} \in \mathbb{F}_p$. Then

$$QP = \sum_{k=0}^{n+d} (r_{k,0} + r_{k,1}\alpha + ... + r_{k,m-1}\alpha^{m-1})x^k.$$

We want QP to be in $\mathcal{F}[x^m]$, where $\mathcal{F}$ is the fixed field of θ, and we assume that $deg(QP) = n + d = nm^2$. For QP to be in $\mathcal{F}[x^m]$ we want:

$$\begin{aligned} r_{k,i} &= 0 && \text{if } m \nmid k \\ r_{\mu m,i} &= 0 && \text{if } i > 0 \end{aligned}$$

In other words, $r_{k,i}$ is only nonzero when m divides k and $i = 0$. Note that $r_{i,j}$ are $\mathbb{F}_p$-linear forms in $q_{i,j}$. To have a non-trivial solution we want the number of variables to be larger than the number of equations. The number of variables is the number of coefficients in Q, which is $(d+1)m$. The number of equations is $(n+d+1)m - nm - 1 = dm + m - 1$. So for QP to be in $\mathcal{F}[x^m]$ we want

$$dm + m - 1 < (d+1)m$$

which clearly holds. We have thus constructed QP, a bound on P, with $deg(QP) = n + d = nm^2$, as desired. □

Next result is due to Solé. He proved a better bound of a polynomial in special case of $\mathbb{F}_4[x, \theta]$.

Lemma 5. *The bound of a polynomial of degree r in $\mathbb{F}_4[x, \theta]$ is of degree at most $2r$.*

Proof. Let

$$g = \sum_{i=0}^{r} g_i x^i$$
$$\tilde{g} = \sum_{i=0}^{r} \theta^{i+1}(g_i) x^i = \sum_{i=0}^{r} x^i \theta(g_i).$$

Then we can compute $g\tilde{g}$ and order the terms:

$$\begin{aligned} g\tilde{g} &= \sum_{i,j}^{r} g_i x^{i+j} \theta(g_i) \\ &= \sum_{k=0}^{2r} \sum_{i+j=k} g_i x^k \theta(g_i) \\ &= \sum_{k=0}^{2r} \sum_{i+j=k} g_i \theta^{k+1}(g_i) x^k. \end{aligned}$$

Let

$$a_k = \sum_{i+j=k} g_i \theta^{k+1}(g_j).$$

Consider now the parity of k. For terms with odd k we use the fact that $k+1$ is even and that we are in $\mathbb{F}_{2^2}$ which implies that every even power of θ is an identity map and $k+1$ terms cancel out:

$$\begin{aligned} a_k &= \sum_{i+j=k} g_i \theta^{k+1}(g_j) \\ &= \sum_{i+j=k} g_i g_j \\ &= g_0 g_k + g_1 g_{k-1} + ... + g_{k-1} g_1 + g_k g_0 \\ &= 2 g_0 g_k + ... + 2 g_{\frac{k-1}{2}} g_{\frac{k+1}{2}} \\ &= 0 \end{aligned}$$

For terms with even k, we have

$$\begin{aligned} a_k &= \sum_{i+j=k} g_i \theta^{k+1}(g_j) \\ &= \sum_{i+j=k} g_i \theta(g_j). \end{aligned}$$

But note that

$$\begin{aligned} \theta(a_k) &= \sum_{i+j=k} \theta^{k+1}(g_i) g_j \\ &= \sum_{i+j=k} \theta(g_i) g_j \\ &= a_k \end{aligned}$$

so $a_k \in \mathbb{F}_2$ and because $a_k = 0$ for odd k we have that $g\tilde{g} \in \mathbb{F}_2[x^2]$. The degree of $g\tilde{g}$ is $2r$ and it generates a two sided ideal. Thus the bound on g is of degree at most $2r$. $\square$

Chapter 3

Factorisation in Commutative Polynomial Rings

Any nonconstant polynomial over a field can be expressed as a product of irreducible polynomials. The problem of finding the irreducible factors can be solved efficiently over finite fields. In this chapter we will look at factorisation in commutative polynomial rings over finite fields as a warm up for the next chapter which is devoted to factorisation in the skew polynomial rings.

There are many algorithms for factorisation in commutative polynomial rings over finite fields. The choice of algorithm depends on the size of the underlying field. We will present here only one algorithm, the Berlekamp's algorithm, that is suitable for "small" fields. For more algorithms for factorisation over finite fields see [10] or [11, Chapter 4].

3.1 Berlekamp's algorithm

A polynomial $f \in \mathbb{F}_q[x]$ can be written as a product of irreducible factors $f_i \in \mathbb{F}_q[x]$

$$f = f_1^{e_1} ... f_r^{e_r}$$

with $e_i \in \mathbb{N}$. We call factors $f_i^{e_i}$ the primary factors of f. This factorisation is unique since $\mathbb{F}_q[x]$ is a unique factorisation domain. Berlekamp's algorithm takes as input a polynomial $f \in \mathbb{F}_q[x]$ and produces as output the primary factors of f. The following lemma is crucial.

Lemma 6. *If $f \in \mathbb{F}_q$ is monic and $v \in \mathbb{F}_q$ is such that $v^q \equiv v \mod f$, then*

$$f = \prod_{a \in \mathbb{F}_q} gcd(f, v - a).$$

Proof. We know that the elements of $\mathbb{F}_q$ are exactly the q distinct zeros of the polynomial $x^q - x$. Then $x^q - x = \prod_{a \in \mathbb{F}_q}(x - a)$. Further, since v is such that $v^q \equiv v$

$\mod f$, $v^q - v$ is a multiple of f and thus $gcd(f, v^q - v) = f$. Then

$$f = gcd(f, v^q - v) = gcd(f, \prod_{a \in \mathbb{F}_q} (v - a)).$$

Note that $gcd(v - a, v - b) = 1$ whenever $a \neq b$. Then $gcd(f, \prod_{a \in \mathbb{F}_q}(v - a)) = \prod_{a \in \mathbb{F}_q} gcd(f, v - a)$ and we are done. □

For this factorisation we need calculations of q greatest common divisors, which is accomplished by application of Euclidean algorithm, but it is efficient only for small finite fields. In general, however, this does not give a factorisation of f into primary factors. The idea is to keep applying this procedure to each of the factors obtained until we find primary factors. Thus we have to construct a method for finding polynomials $v \in \mathbb{F}_q[x]$ such that $v^q \equiv v \mod f$. For this we will need to make use of the ring isomorphism

$$\mathbb{F}_q[x]/(f) \simeq \oplus_{i=1}^{r} \mathbb{F}_q[x]/(f_i^{e_i}).$$

implied by the Chinese remainder theorem. For the rest of the discussion we will restrict our attention to factoring monic polynomials with no repeated factors. This is not really a restriction because the problem of factoring a polynomial in $\mathbb{F}_q[x]$ can be reduced to a problem of factoring a number of monic, squarefree polynomials in the following way.

Compute the greatest common divisor of $f(x)$ and its derivative

$$d(x) = gcd(f(x), f'(x)).$$

We distinguish between three cases.

I If $d(x) = 1$ then $f(x)$ has no repeated factors.

II If $d(x) = f(x)$ then $f'(x) = 0$. This implies that $f(x) = g(x)^p$ for some $g(x)$ in $\mathbb{F}_q[x]$ and p the characteristic of $\mathbb{F}_q$. It is not necessarily the case that $g(x)$ is squarefree, but we can repeat the reduction procedure on $g(x)$ until we get a squarefree polynomial.

III If neither of the two cases above holds and $d(x)$ contains common factors, then $f(x)/d(x)$ is squarefree. To factor $f(x)$ in this case, we factor $d(x)$ and $f(x)/d(x)$ separately. It can of course happen that $d(x)$ is not squarefree, in which case we apply this reduction on $d(x)$ until it is squarefree.

We can now state the following theorem assuming that $f(x)$ is squarefree.

Theorem 9 (cf. page 131 in [11])**.** *Set of solutions of $v^q \equiv v \mod f$ is a $\mathbb{F}_q$-linear vector space of dimension r.*

Proof. Let $f = f_1 ... f_r$ be a product of distinct monic irreducible factors over $\mathbb{F}_q$, let $(a_1, ..., a_r)$ be any r-tuple with $a_i \in \mathbb{F}_q$ and let v be a solution to $v^q \equiv v \mod f$. Then from

$$v^q - v = \prod_{a \in \mathbb{F}_q} (v - a)$$

we see that every irreducible factor of f divides $v - a$ for some $a \in \mathbb{F}_q$. Then for all solutions v of $v^q \equiv v \mod f$ it must be that $v \equiv a_i \mod f_i$ for some r-tuple $(a_1, ..., a_r)$ over $\mathbb{F}_q$.

On the other hand, by the Chinese remainder theorem we know there is a unique $v \in \mathbb{F}_q[x]$ that satisfies

$$v \equiv a_i \mod f_i \text{ with } 1 \leqslant i \leqslant r.$$

Then $v^q \equiv a_i^q = a_i \equiv v \mod f_i$ and there are exactly q^r solutions over $\mathbb{F}_q$. □

For the discussion on cyclic codes it is interesting to look at the factorisation of a special polynomial, $x^n - 1$. We will now look at its decomposition in $\mathbb{F}_q[x]$. Later, in section 4.3 on factorisation in skew rings, we will come back to polynomial $x^n - 1$ and consider its decomposition in $\mathbb{F}_q[x, \theta]$.

3.2 Factorisation of $x^n - 1$

Definition 15. The nth cyclotomic polynomial Φ_n, for any positive integer n, is defined as

$$\Phi_n(x) = \prod_{\omega} (x - \omega)$$

where the product is over all nth primitive roots of unity ω over a field. It is clear that $\Phi_n(x)$ has degree $\phi(n)$, the Euler's totient function of n.

Proposition 2.

$$x^n - 1 = \prod_{d|n} \Phi_d(x)$$

This is a standard result. For a proof see for example Proposition 9.1.5 in [4].

Remark 1. Consider n and q with $gcd(n, q) \neq 1$. Let $q = p^e$. Then we can write $n = mp^s$ with positive integers m and s and $gcd(p, m) = 1$. Then the factors of $x^n - 1$ are factors of $x^m - 1$, each occurring with multiplicity p^s. This is not hard to see:

$$x^n - 1 = x^{mp^s} - 1 = (x^m - 1)^{p^s}.$$

It follows that it is sufficient to consider the decomposition in case when q and n are relatively prime.

Note that Frobenius action partitions $\mathbb{Z}/n\mathbb{Z}$ into cyclotomic cosets

$$C_s = (s, sq, sq^2, ...sq^{m_s 1})$$

where $m_s = |C_s|$ and $sq^{ms} \equiv s \mod n$. For example, let $q = 5$ and $n = 9$. Then $\mathbb{Z}/9\mathbb{Z} = \{0\} \cup \{3, 6\} \cup \{1, 5, 7, 8, 4, 2\}$. Each cyclotomic coset corresponds to a factor

of $x^n - 1$ with coefficients in $\mathbb{F}_q$, irreducible over $\mathbb{F}_q[x]$. To see why this is the case note that

$$\begin{aligned} x^n - 1 &= \prod_{i=1}^{n} (x - \alpha^i),\ \alpha \in \overline{\mathbb{F}}_q \\ &= \prod_{s} \prod_{i \in C_s} (x - \alpha^i) \end{aligned}$$

where the product of the second equality is over all cosets C_s. In the example above, if $\alpha \in \overline{\mathbb{F}}_q$ is a root of a polynomial then so are $\alpha^5, \alpha^7, \alpha^8, \alpha^4$ and α^2. Thus the length of a coset is the degree of the corresponding irreducible factor. So in the same example we have

$$x^9 - 1 = \underbrace{(x - \alpha^0)}_{x-1} \underbrace{(x - \alpha^3)(x - \alpha^6)}_{x^2+\ldots} \underbrace{(x - \alpha)(x - \alpha^5)(x - \alpha^7)(x - \alpha^8)(x - \alpha^4)(x - \alpha^2)}_{x^6+\ldots}$$

over $\overline{\mathbb{F}}_5[x]$ or

$$x^9 - 1 = (x - 1)(x^2 + \ldots)(x^6 + \ldots)$$

over $\mathbb{F}_5[x]$ or

$$x^9 - 1 = \Phi_1(x)\Phi_3(x)\Phi_9(x)$$

by proposition 2.

Cyclotomic polynomials are not necessarily irreducible in $\mathbb{F}_q[x]$. Let $q = 3$ and $n = 8$. By proposition 2 we have $x^8 - 1 = \Phi_1(x)\Phi_2(x)\Phi_4(x)\Phi_8(x)$. However, $\mathbb{Z}/8\mathbb{Z} = \{0\} \cup \{1, 3\} \cup \{2, 6\} \cup \{4\} \cup \{5, 7\}$. Then

$$x^8 - 1 = \underbrace{(x - \alpha^0)}_{x-1} \underbrace{(x - \alpha)(x - \alpha^3)}_{x^2+\ldots} \underbrace{(x - \alpha^2)(x - \alpha^6)}_{x^2+\ldots} \underbrace{(x - \alpha^4)}_{x+\ldots} \underbrace{(x - \alpha^5)(x - \alpha^7)}_{x^2+\ldots}$$

and so

$$x^8 - 1 = \underbrace{(x - 1)}_{\Phi_1(x)} \underbrace{(x + \ldots)}_{\Phi_2(x)} \underbrace{(x^2 + \ldots)}_{\Phi_4(x)} \underbrace{(x^2 + \ldots)(x^2 + \ldots)}_{\Phi_8(x)}$$

over $\mathbb{F}_3[x]$.

Chapter 4

Factorisation in Skew Polynomial Rings

In commutative polynomial rings we saw that Berlekamp's algorithm finds factors of any polynomial $f \in \mathbb{F}_q[x]$. This chapter will be devoted to factorisation in skew polynomial rings $\mathbb{F}_q[x, \theta]$. An efficient algorithm for factorisation already exists and is given by Giesbrecht in [6]. It is rather complicated and hard to follow and we do not present it here but curious readers are encouraged to read his paper. Instead, we present another approach to factorisation problem that uses linear difference operators. In preparing this chapter I relied heavily on [1].

4.1 Difference Operators

Let p be a prime, $q = p^m$ some power of p and θ the Frobenius map on $\mathbb{F}_q$.

Definition 16. An $\mathbb{F}_p$-linear operator is an $\mathbb{F}_p$-linear map from a vector space to itself.

Then we can say that θ is an $\mathbb{F}_p$-linear operator on $\overline{\mathbb{F}}_p$, the algebraic closure of $\mathbb{F}_p$.

$$\begin{aligned} \theta : \overline{\mathbb{F}}_p &\to \overline{\mathbb{F}}_p \\ \alpha &\mapsto \alpha^p. \end{aligned}$$

Also note that we can look at any element a of $\mathbb{F}_q$ as a linear operator in the following way

$$\begin{aligned} a : \overline{\mathbb{F}}_p &\to \overline{\mathbb{F}}_p \\ \alpha &\mapsto a\alpha. \end{aligned}$$

Since any linear combination of $\mathbb{F}_p$-linear operator θ and any $a \in \mathbb{F}_q$ is also an $\mathbb{F}_p$-linear operator we can define the set of all $\mathbb{F}_p$-linear operators generated by them:

$$\mathbb{F}_q[\theta] = \{a_0 + a_1\theta + ... + a_n\theta^n : a_i \in \mathbb{F}_q\}.$$

With addition the usual addition of polynomials in θ and multiplication the skew multiplication of polynomials in θ this set has a ring structure. In what follows we consider some of its properties.

Definition 17. To any nontrivial element $L \in \mathbb{F}_q[\theta]$ we define the solution space of L to be

$$\mathrm{Sol}(L) = \{\beta \in \overline{\mathbb{F}}_p : L(\beta) = 0\}.$$

Proposition 3. *Let $L = \theta^d + a_1\theta^{d-1} + ... + a_d \in \mathbb{F}_q[\theta]$ be such that $a_d \neq 0$. Then $\mathrm{Sol}(L)$ is an $\mathbb{F}_p$-vector space whose dimension equals the degree of L in θ.*

Proof. First note that $\mathrm{Sol}(L) \subset \overline{\mathbb{F}}_p$ is an $\mathbb{F}_p$-vector space. To see that we want $0 \in \mathrm{Sol}(L)$, i.e. $L(0) = 0$, which is indeed the case:

$$L(0) = (\theta^d + a_1\theta^{d-1} + ... + a_d)(0) = 0.$$

Next, for any $\alpha, \beta \in \mathrm{Sol}(L)$ we have that $\alpha + \beta \in \mathrm{Sol}(L)$. The crucial observation here is that $(\alpha + \beta)^p = \alpha^p + \beta^p$.

$$\begin{aligned}
L(\alpha + \beta) &= (\theta^d + a_1\theta^{d-1} + ... + a_d)(\alpha + \beta) \\
&= \theta^d(\alpha + \beta) + a_1\theta^{d-1}(\alpha + \beta) + ... + a_d(\alpha + \beta) \\
&= (\alpha + \beta)^{p^d} + a_1(\alpha + \beta)^{p^{d-1}} + ... + a_d(\alpha + \beta) \\
&= \alpha^{p^d} + \beta^{p^d} + a_1\alpha^{p^{d-1}} + a_1\beta^{p^{d-1}} + ... + a_d\alpha + a_d\beta \\
&= \theta^d(\alpha) + \theta^d(\beta) + a_1\theta^{d-1}(\alpha) + a_1\theta^{d-1}(\beta) + ... + a_d\alpha + a_d\beta \\
&= L(\alpha) + L(\beta) \\
&= 0.
\end{aligned}$$

For any $\alpha \in \mathrm{Sol}(L)$ and any $\lambda \in \mathbb{F}_p$, $\lambda\alpha \in \mathrm{Sol}(L)$ holds because $\theta^i(\lambda) = \lambda$ for any i since $\mathbb{F}_p$ is a fixed field of θ, and so we can just take λ in front $L(\lambda\alpha) = \lambda L(\alpha) = 0$.

We need to show next that the dimension of $\mathrm{Sol}(L)$ equals the degree of L in θ. Let $deg(L) = d$ and write

$$L = \theta^d + a_1\theta^{d-1} + ... + a_d$$

where $a_d \neq 0$. Now we want to count the number of solutions to L:

$$\begin{aligned}
|\mathrm{Sol}(L)| &= \#\{\beta : L(\beta) = 0\} \\
&= \#\{\beta : \beta^{p^d} + a_1\beta^{p^{d-1}} + ... + a_{d-1}\beta^p + a_d\beta = 0\}.
\end{aligned}$$

The polynomial in β is of degree p^d so there are exactly p^d solutions if they are all distinct. In that case the vector space $\mathrm{Sol}(L)$ with p^d distinct points has dimension d over $\mathbb{F}_p$. So we only need to see that all solutions are distinct. Note that $\beta = 0$ is one solution and write

$$\beta = 0 \vee \underbrace{\beta^{p^d-1} + a_1\beta^{p^{d-1}-1} + ... + a_{d-1}\beta^{p-1} + a_d}_{F(x)} = 0.$$

None of the solutions of $F(x) = 0$ are equal to zero since $a_d \neq 0$. Further, $F(x)$ has distinct solutions if and only if $gcd(F(x), F'(x)) = 1$. Then write

$$\begin{aligned} F(x) &= x^{p^d-1} + a_1 x^{p^{d-1}-1} + ... + a_{d-1}x^{p-1} + a_d \\ xF'(x) &= (p^d-1)x^{p^d-1} + a_1(p^{d-1}-1)x^{p^{d-1}-1} + ... + a_{d-1}(p-1)x^{p-1} \\ &= -x^{p^d-1} - a_1 x^{p^{d-1}-1} - ... - a_{d-1}x^{p-1}. \end{aligned}$$

The sum of $F(x)$ and $xF'(x)$ now very nicely simplifies to a_d and we are done, since if $F(x)$ and $xF'(x)$ had a factor in common it would divide $a_d \neq 0$. □

Remark 2. In fact, it could be shown that if we don't require $a_d \neq 0$ that $\mathrm{Sol}(L)$ is an $\mathbb{F}_p$-vector space whose dimension is at most equal to $deg_\theta(L)$. More precisely, if $L = a_n\theta^n + ... + a_0$ with $a_k \neq 0$ and $a_i = 0$ for all $i < k$ then $dim(\mathrm{Sol}(L)) = n - k$. To see this note that $L = L_1\theta^k$ for some $L_1 \in \mathbb{F}_q[\theta]$ with $deg_\theta(L_1) = n - k$. Since multiplication in $\mathbb{F}_q[\theta]$ is just a composition of operators, saying that β is a solution of L, $L(\beta) = 0$, is the same as saying $L_1(\theta(\beta)) = L_1(\beta^{p^k}) = 0$. By a counting argument like above we see that $\mathrm{Sol}(L)$ is a vector space of dimenstion $deg_\theta(L_1) = n - k$.

Proposition 4. *Let $L \in \mathbb{F}_q[\theta]$ with $q = p^m$. Then $\mathrm{Sol}(L)$ is stable under θ^m.*

Proof. To show that $\mathrm{Sol}(L)$ is θ^m-stable we need to see that $\theta^m(\beta) \in \mathrm{Sol}(L)$ for all $\beta \in \mathrm{Sol}(L)$ where

$$\begin{aligned} \theta^m : \overline{\mathbb{F}}_p &\to \overline{\mathbb{F}}_p \\ \alpha &\mapsto \alpha^{p^m}. \end{aligned}$$

Let $\beta \in \mathrm{Sol}(L)$ and $L = \theta^d + a_1\theta^{d-1} + ... + a_d$. By definition $L(\beta) = \beta^{p^d} + a_1\beta^{p^{d-1}} + ... + a_d\beta = 0$. Let θ^m act on β:

$$\theta^m(\beta) = \beta^{p^m} = \beta^q.$$

We need to show that for every $\beta \in \mathrm{Sol}(L)$ it is also the case that $\beta^q \in \mathrm{Sol}(L)$:

$$\begin{aligned} L(\beta^q) &= (\theta^d + a_1\theta^{d-1} + ... + a_d)(\beta^q) \\ &= \theta^d(\beta^q) + a_1\theta^{d-1}(\beta^q) + ... + a_d\beta^q \\ &= \beta^{q\cdot p^d} + a_1\beta^{q\cdot p^{d-1}} + ... + a_d\beta^q \\ &= (\beta^{p^d} + a_1\beta^{q\cdot p^{d-1}} + ... + a_d)^q \\ &= L(\beta)^q \\ &= 0. \end{aligned}$$

□

Proposition 5. *To any θ^m-stable $\mathbb{F}_p$-vector space $V \subset \overline{\mathbb{F}}_p$ of dimension n there corresponds a unique monic element $L \in \mathbb{F}_q[\theta]$ of degree n such that $\mathrm{Sol}(L) = V$. Denote such L by L_V.*

Proof. Let V be a θ^m-stable $\mathbb{F}_p$-vector space of dimension n and let $\alpha_1, ..., \alpha_n$ be its basis. Consider the action of θ on the basis:

$$\begin{gathered}\alpha_1, ..., \alpha_n \\ \theta(\alpha_1), ..., \theta(\alpha_n) \\ \vdots \\ \theta^n(\alpha_1), ..., \theta^n(\alpha_n).\end{gathered}$$

We see that α_i has $n+1$ conjugates and so there must exist $\lambda_0, ..., \lambda_n \in \overline{\mathbb{F}}_p$, not all zero, such that

$$\lambda_n\theta^n(\alpha_i) + ... + \lambda_1\theta(\alpha_i) + \lambda_0\alpha_i = 0 \text{ for } i = 1, ..., n.$$

So

$$\lambda_n\theta^n(\beta) + ... + \lambda_1\theta(\beta) + \lambda_0\beta = 0 \text{ for all } \beta \in V. \qquad (*)$$

Now we have $\lambda_i \in \overline{\mathbb{F}}_p$ for all i, but we would like a polynomial with coefficients in $\mathbb{F}_q$. We know $\overline{\mathbb{F}}_p$ is infinite, but we can let $\mathbb{F}_{q^N}$ denote a large enough field extension of $\mathbb{F}_q$ such that for all i, $\lambda_i \in \mathbb{F}_{q^N}$. Then if we apply powers of θ on $(*)$ we get

$$\lambda_n^{q^r}\theta^n(\beta) + ... + \lambda_1^{q^r}\theta(\beta) + ... + \lambda_0^{q^r}\beta = 0$$

for all $\beta \in V$ and all $r = 0, ..., N-1$. Note that the action of θ only "twists" the coefficients and permutes β's since V is θ^m-stable. If we take the trace of λ_i in $\mathbb{F}_{p^N}$ and denote it by μ_i

$$\mu_i = Tr(\lambda_i) = \sum_{r=0}^{N-1} \lambda_i^{q^r} \in \mathbb{F}_q$$

we get an element of $\mathbb{F}_q$. Then the polynomial

$$\mu_n\theta^n(\beta) + ... + \mu_1\theta(\beta) + ... + \mu_0\beta \in \mathbb{F}_q[\theta]$$

evaluates to zero for all $\beta \in V$ and we would be done except that it could be that all μ_i's are zero. We know however that there exists $\lambda \in \mathbb{F}_{q^N}$ such that its trace $\mu = \sum_{r=0}^{N-1} \lambda^{q^r}$ is non-zero. Then take a non-zero λ_s and normalize it to λ by taking the operators

$$\frac{\lambda}{\lambda_s}(\lambda_n\theta^n + ... + \lambda_1\theta + \lambda_0).$$

Then we have a non-trivial polynomial which annihilates V with $dim(V) = n$, thus it's degree is n. □

Proposition 6. *For any two θ^m-stable $\mathbb{F}_p$-vector spaces U, V we have*

(i) $L_{U\cap V} = gcrd(L_U, L_V)$

(ii) $L_{U+V} = lclm(L_U, L_V)$

Proof.

(i) Let $g = gcrd(L_U, L_V)$. This means that

$$L_U = hg \text{ for some } h \in \mathbb{F}_q[\theta]$$
$$L_V = h'g \text{ for some } h' \in \mathbb{F}_q[\theta]$$

and thus $\mathrm{Sol}(L_U) \cap \mathrm{Sol}(L_V) \supset \mathrm{Sol}(g)$. For the other direction note that the Bézout's identity and existence of the extended Euclidean algorithm implies that

$$\exists A, B \in \mathbb{F}_q[\theta] \text{ such that } AL_U + BL_V = g.$$

Consider $\alpha \in \mathrm{Sol}(L_U) \cap \mathrm{Sol}(L_V)$, i.e. $L_U(\alpha) = 0$ and $L_V(\alpha) = 0$. Then evaluated at α:

$$AL_U(\alpha) + BL_V(\alpha) = g(\alpha) \text{ for same } A, B \in \mathbb{F}_q[\theta].$$

Obviously $g(\alpha) = 0$ and it follows that $\mathrm{Sol}(L_U) \cap \mathrm{Sol}(L_V) \subset \mathrm{Sol}(g)$. So $\mathrm{Sol}(L_U) \cap \mathrm{Sol}(L_V) = \mathrm{Sol}(g)$. From the correspondence given in the Proposition 5 we know that $\mathrm{Sol}(L_U) = U$, $\mathrm{Sol}(L_V) = V$ and thus $U \cap V = \mathrm{Sol}(g)$. Also, $\mathrm{Sol}(L_{U\cap V}) = U \cap V$ and thus $\mathrm{Sol}(L_{U\cap V}) = \mathrm{Sol}(g)$. We finally conclude that $L_{U\cap V} = g$.

(ii) The proof for this statement is similar. One just needs to notice that if we let $g = lclm(L_U, L_V)$, then $\mathrm{Sol}(L_U) + \mathrm{Sol}(L_V) = \mathrm{Sol}(g)$, where $\mathrm{Sol}(L_U) + \mathrm{Sol}(L_V) \supset \mathrm{Sol}(g)$ follows immediately, and $\mathrm{Sol}(L_U) + \mathrm{Sol}(L_V) \subset \mathrm{Sol}(g)$ follows from the Bézout's identity. □

Proposition 7. *Denote by L_U and L_V the polynomials in $\mathbb{F}_q[\theta]$ whose solution spaces are U and V respectively. Then V is a proper, θ^m-stable vector subspace of U if and only if L_V is a right factor of L_U.*

Proof. By Proposition 6 (i) and $V \subset U$ we have that

$$L_V = L_{U\cap V} = gcrd(L_U, L_V).$$

By definition of $gcrd$ there exists a polynomial M in $\mathbb{F}_q[\theta]$ such that

$$L_U = ML_V$$

and thus L_V is a right factor of L_U. The converse is similar: If L_V is a right factor of L_U then $L_V = gcrd(L_U, L_V)$ by definition of $gcrd$. By Proposition 6 (i) $L_{U\cap V} = gcrd(L_U, L_V)$ and so $L_V = L_{U\cap V}$ and thus $V \subset U$. □

Remark 3. Note that U is irreducible under the action of θ^m if and only if L_U is irreducible. This is equivalent to the statement that U is reducible if and only if L_U is reducible, which follows directly from Proposition 7.

4.2 Cayley-Hamilton and Implications

Let g be a polynomial in $\mathbb{F}_q[x, \theta]$. Denote by (g) the left ideal generated by left multiples of g. Recall the the bound on g was defined as a generator g^* of the maximal

two-sided ideal (g^*) that is contained in (g). We also saw that g^* must be of the form $x^t h$ where t is some integer and h is a central element of $\mathbb{F}_q[x,\theta]$.

Using the terminology of the ring of operators $\mathbb{F}_q[\theta]$, we say that the bound of $L \in \mathbb{F}_q[\theta]$ is an operator $x^t K$ with t an integer and $K \in \mathcal{F}[\theta^m]$ of smallest degree such that it is divisible by L, where $\mathcal{F}$ is the fixed field of θ. For the remainder of this text I will consider a special case when θ is the Frobenius automorphism ($\mathcal{F} = \mathbb{F}_p$). Generalization to powers of Frobenius is straightforward.

Proposition 8. *Let θ be the Frobenius automorphism of $\mathbb{F}_q$ with $q = p^m$. If $L \in \mathbb{F}_q[\theta]$ is of degree d then its bound is of degree at most md.*

Proof. Let $L = \theta^d + a_1\theta^{d-1} + ... + a_d$. From Proposition 3 we know that $\mathrm{Sol}(L)$ is an $\mathbb{F}_p$-vector space whose dimension is $deg_\theta(L) \leqslant d$. Let $\alpha_1, ..., \alpha_d$ be $\mathbb{F}_p$-basis of $\mathrm{Sol}(L)$. From Proposition 4 we know $\mathrm{Sol}(L)$ is stable under θ^m. Since θ^m is an $\mathbb{F}_p$-linear map

$$\theta^m : \mathrm{Sol}(L) \to \mathrm{Sol}(L)$$

we can represent it by a matrix T that acts on basis

$$\theta^m \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{pmatrix} = \underbrace{\begin{pmatrix} t_{11} & ... & t_{1d} \\ \vdots & & \vdots \\ t_{d1} & ... & t_{dd} \end{pmatrix}}_{T} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{pmatrix}$$

with $t_{ij} \in \mathbb{F}_p$. A theorem by Cayley and Hamilton states that every square matrix over a commutative ring satisfies its own characteristic equation. In our case $\mathbb{F}_p$-linear map θ^m is represented by a square matrix T over $\mathbb{F}_p$ whose characteristic polynomial $det(T - \lambda I_d)$ we denote by $f_L(\lambda)$. By Cayley-Hamilton $f_L(T) = 0$, or in terms of θ^m

$$f_L(\theta^m) = 0 \text{ on } \mathrm{Sol}(L).$$

Note that $f_L(\theta^m)$ is an element of degree d in θ^m of $\mathbb{F}_p[\theta^m]$, the center of $\mathbb{F}_q[\theta]$, and so it generates a two-sided ideal which is contained in left ideal generated by L. The latter is true since if L was not a right factor of $f_L(\theta^m)$ we could write

$$f_L(\theta^m) = AL + R$$

and R with $deg(R) < deg(L) = d$ would annihilate $\mathrm{Sol}(L)$ which is impossible since $\mathrm{Sol}(L)$ is of dimension d. Thus $R = 0$. Finally, the bound of L is of degree at most $deg_\theta(f_L(\theta^m)) = md$. □

Boucher and Ulmer derived the same bound to be at most $m^2 d$. We showed here that the bound for the special case of $\mathbb{F}_4[x,\theta]$ Sóle derived ($\leqslant md$), is in fact still the case for general $\mathbb{F}_q[x,\theta]$.

The correspondence between elements of $\mathbb{F}_q[\theta]$ and θ^m-stable $\mathbb{F}_p$-linear vector spaces as described in Proposition 5 and characteristic polynomials as a consequence of Cayley-Hamilton theorem, as we understand it so far, is described in the diagram below.

$$\begin{array}{lcll} L_U & = M \cdot & L_V & \in \mathbb{F}_q[\theta] \\ \Big\downarrow \text{Sol} & & \Big\downarrow \text{Sol} & \\ U & \supseteq & V & \\ \Big\downarrow CH & & \Big\downarrow CH & \\ f_U & & f_V & \in \mathbb{F}_p[\theta^m] \end{array}$$

Let L_V be an irreducible right factor of L_U of degree r in $\mathbb{F}_q[\theta]$. Let U and V denote $\mathrm{Sol}(L_U)$ and $\mathrm{Sol}(L_V)$ respectively. We know V is an irreducible θ^m-stable subspace of U (see proof of Proposition 3) of dimension r. Let f_U denote a characteristic polynomial of $\mathbb{F}_p$-linear map $\theta^m : U \to U$ and f_V a characteristic polynomial of $\mathbb{F}_p$-linear map $\theta^m : V \to V$. In previous section we saw that f_U and f_V are polynomials in $\mathbb{F}_p[\theta^m]$, the center of $\mathbb{F}_q[\theta]$, and by Cayley-Hamilton $f_U(\theta^m) = 0$ on U and $f_V(\theta^m) = 0$ on V. Note that degree of f_V in θ^m is r.

Proposition 9. *Let notation be as above. Then f_V is irreducible and it divides f_U.*

Proof. Suppose $f_V = h_1 h_2$ with $deg(h_1) = r_1$, $deg(h_2) = r_2$ and $0 < r_1, r_2 < r$. We know by Cayley-Hamilton that

$$f_V(\theta^m)(\beta) = 0 \text{ for all } \beta \in V.$$

From the decomposition of f_V it follows that

$$h_1(\theta^m)h_2(\theta^m)(\beta) = 0 \text{ for all } \beta \in V.$$

Consider $\{h_2(\theta^m)(\beta) : \beta \in V\}$, a θ^m-stable subspace of V. But V is irreducible and so $\{h_2(\theta^m)(\beta) : \beta \in V\}$ is either equal to V or zero. We will show now that both cases lead to contradiction. Let $h_2(\theta^m)(\beta) = 0$ for all $\beta \in V$. This implies that $span(\beta, \theta^m(\beta), ..., (\theta^m)^{r_2-1}(\beta))$ is θ^m-stable while of dimension $r_2 < r$. Contradiction. If on the other hand $\{h_2(\theta^m)(\beta) : \beta \in V\} = V$, from decomposition of f_V we get

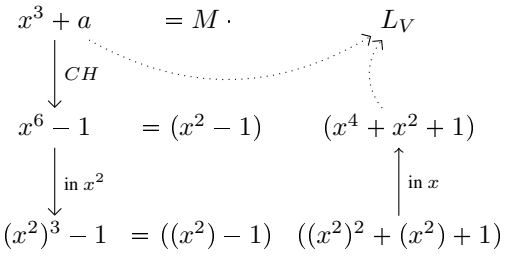
$$0 = f_V(\theta^m)(\beta) = h_1(\theta^m)h_2(\theta^m)(\beta) \text{ for all } \beta \in V.$$

So $h_1(\theta^m)(\beta) = 0$ for all $\beta \in V$. We can repeat the same argument as for h_2 to reach contradiction. □

This lemma allows us to update the diagram as follows

$$\begin{array}{lcll} L_U & = M \cdot & L_V & Skew \\ \Big\downarrow & & \Big\downarrow & \\ f_U & = g \cdot & f_V & Commutative \end{array}$$

with L_V and f_V irreducible, and suggests an idea for a factorisation algorithm. We are given a polynomial $L_U \in \mathbb{F}_q[x, \theta]$ for which we want to find a right irreducible factor, L_V. We first compute a bound f_U on L_U as described in section 4.2. The bound f_U is a polynomials in x^m where m is the order of θ. We then factor it into irreducible factors in x^m using Berlekamp's algorithm, since we are now in the center of skew ring, which is commutative. Pick one irreducible factor of f_U and call it f_V. We know by construction in section 4.2 and proposition 9 that $deg_{x^m}(f_V) = deg(L_V)$. The problem that remains is how to find L_V. One idea is to compute the $gcrd$ of $f_V(x)$ and L_U. If the result is of degree $deg_{x^m}(f_V)$ then we have found L_V. If not, then we know that L_V is contained in $gcrd(f_V(x), L_U)$. Diagram below gives an example of this algorithm in $\mathbb{F}_4[x, \theta]$.



In the example given in the diagram we are lucky because

$$gcrd(x^3 + a,\ x^4 + x^2 + 1) = x^2 + a^2x + 1$$

which is of degree 2, thus irreducible in $\mathbb{F}_q[\theta]$, and we have found a right irreducible factor of L_U. In case we are not as lucky as in the example above, the $gcrd(L_U, f_V)$ is reducible. But even when that is the case, we can still compute L_V. Namely, L_V must divide $gcrd(L_U, f_V)$ and since $deg(gcrd(L_U, f_V)) < deg(L_U)$ we have reduced the problem to finding a right irreducible factor of a polynomial of a strictly smaller degree. We can then repeat the same procedure until we get

$$deg_{\theta^m}(f_V) = deg(gcrd(L_U, f_V)).$$

4.3 Factorisation of $x^n - 1$

The idea behind studying the decomposition of $x^n - 1$ in commutative polynomial ring was that it would give some insight into its decomposition in skew polynomial ring. These rings have a very different structure and it doesn't come as a surprise that they in fact don't have similar decompositions of arbitrary polynomials. For one, skew polynomial ring is not a UFD. Furthermore, a polynomial that is irreducible in $\mathbb{F}_q[x]$ might very well be reducible in $\mathbb{F}_q[x, \theta]$, and other way around. However, $x^n - 1$ is a rather special polynomial and we are able to predict the degrees of irreducible factors using the theory developed so far (in which we will rely the most on the properties of the characteristic polynomial f_L) and the knowledge of commutative factorisation of $x^n - 1$ in $\mathbb{F}_p[x]$.

For some special n's we can immediately predict the degrees of irreducible factors, as the following example shows, but for a general case we will have to develop a bit more machinery.

Example 5. *When n is a power of p then $x^n - 1$ factors into n linear terms simply because*

$$x^n - 1 = (x-1)^n.$$

Since $\mathbb{F}_q[x,\theta]$ is not a UFD, this is not the only factorisation. Every other factorisation, however, will have n linear terms by theorem 10. Consider $\mathbb{F}_4[x,\theta]$, and denote by a the generator of $\mathbb{F}_4$. These are some of the possible factorisations of $x^4 - 1$:

$$\begin{aligned} x^4 - 1 &= (x-1)^4 \\ &= (ax+1)(ax+a)(x+1)(x+a^2) \\ &= (x+a^2)^2(x+a)^2. \end{aligned}$$

Next easy observation is that when n is a multiple m, the order of θ, we have

$$x^n - 1 = x^{mk} - 1 = (x^m)^k - 1.$$

We know that a polynomial of this form is in center of $\mathbb{F}_q[x,\theta]$, which is commutative. Hence we can use the factorisation of $x^k - 1$ in $\mathbb{F}_q[x]$ to obtain a factorisation of $(x^m)^k - 1$ in x^m. In $\mathbb{F}_4[x,\theta]$ we have for example

$$x^6 - 1 = (x^2)^3 - 1.$$

We know $x^3 - 1 = (x+1)(x+a)(x+a^2)$ in $\mathbb{F}_4[x]$, from which it follows that

$$\begin{aligned} x^6 - 1 &= (x^2)^3 - 1 \\ &= (x^2+1)(x^2+a)(x^2+a^2). \end{aligned}$$

Furthermore note that $x^2 + 1$ is again a polynomial of form $x^n - 1$ in $\mathbb{F}_4[x,\theta]$ with n a multiple of order of θ so we can apply the same argument once again to obtain

$$x^6 - 1 = (x+1)(x+1)(x^2+a)(x^2+a^2)$$

as a factorisation into irreducible factors. There are of course many other possible factorisations. Here are two:

$$\begin{aligned} x^6 + 1 &= (x^2+a)(x^2+a+1)(x+a+1)(x+a) \\ &= (x^2+x+1)^2(x+a)(x+a+1). \end{aligned}$$

By theorem 10 every skew factorisation of $x^6 - 1$ in $\mathbb{F}_4[x,\theta]$ will have four irreducible factors: two of degree two and two of degree one.

When n is neither of these special cases consider $L = x^n - 1 \in \mathbb{F}_p[\theta]$ with $\mathbb{F}_p$-linear map

$$\theta^m : \mathrm{Sol}(L) \to \mathrm{Sol}(L).$$

Denote its characteristic polynomial by f_{L,θ^m}. Since the coefficients of L are in $\mathbb{F}_p$ we have an additional advantage of $\mathrm{Sol}(L)$ being stable under not only θ^m but also under θ. This is interesting because for a polynomial L of this special form it is easy to write down the characteristic polynomial $f_{L,\theta}$ of the map

$$\theta : \mathrm{Sol}(L) \to \mathrm{Sol}(L).$$

By Cayley-Hamilton

$$f_{L,\theta} : \mathrm{Sol}(L) \to 0$$

and trivially

$$L : \mathrm{Sol}(L) \to 0.$$

Thus L divides $f_{L,\theta}$. On the other hand, $deg_\theta(L) = n$ and as already discussed in section 4.2, $deg_\theta(f_{L,\theta}) = n$. We conclude $f_{L,\theta} = L = x^n - 1$. To see the connection between f_{L,θ^m} and $f_{L,\theta}$ let ζ_m denote mth root of unity, $d = (n, m), m' = \frac{m}{d}, n' = \frac{n}{d}$ and write:

$$\begin{aligned}
f_{L,\theta^m}(x) &= det(\theta^m - xI) \\
&= \prod_{i=0}^{m-1} det(\theta - x^{\frac{1}{m}}\zeta_m^i I) \\
&= \prod_{i=0}^{m-1} f_{L,\theta}(x^{\frac{1}{m}}\zeta_m^i) \\
&= \prod_{i=0}^{m-1} (x^{\frac{1}{m}}\zeta_m^i)^n - 1 \text{ in } \mathbb{F}_p[x] \\
&= \prod_{i=0}^{m-1} (x^{\frac{1}{m'}}\zeta_{m'}^i)^{n'} - 1 \\
&= \prod_{i=0}^{m'-1} (x^{\frac{n'}{m'}}\zeta_{m'}^{n'i} - 1)^d \\
&= (x^{n'} - 1)^d \text{ in } \mathbb{F}_p[x].
\end{aligned}$$

This is the key observation with which the problem of predicting the degrees of irreducible factors of $x^n - 1$ in $\mathbb{F}_q[x, \theta]$ is solved. In a special case of $n = mk$ this simply says that $x^n - 1$ factors in $\mathbb{F}_q[x, \theta]$ into factors of same degree as factors of $(x^k - 1)^m$ in $\mathbb{F}_p[x]$. Note as well that if n and m have no factors in common then $x^n - 1$ factors in $\mathbb{F}_q[x, \theta]$ into factors of same degree as factors of $x^n - 1$ in $\mathbb{F}_p[x]$.

The following examples illustrate this result and close the section.

Example 6. *Let $n = 5$ and consider factorisation in $\mathbb{F}_{2^m}[x, \theta]$. Take $m = 2$ for start.*

To see how x^5-1 factors over $\mathbb{F}_4[x,\theta]$ write

$$\begin{aligned} x^5-1 &= f_{x^5-1,\theta^2}(x) \\ &= \prod_{i=0}^{1} f_{x^5-1,\theta}(x^{\frac{1}{2}}\zeta_2^i) \\ &= \prod_{i=0}^{1}[(x^{\frac{1}{2}}\zeta_2^i)^5-1] \\ &= (x^{\frac{5}{2}}-1)^2 \\ &= x^5-1 \textit{ in } \mathbb{F}_2[x] \\ &= (x+1)(x^4+x^3+x^2+x+1) \end{aligned}$$

For $m=3$ and $m=4$ the factorisation is the same: x^5-1 factors in a unique way into the irreducible factors of the same degree as factors of x^5-1 in $\mathbb{F}_2[x]$. In $\mathbb{F}_{2^5}[x,\theta]$ on the other hand we have

$$\begin{aligned} x^5-1 &= \prod_{i=0}^{4}[(x^{\frac{1}{5}}\zeta_5^i)^5-1] \\ &= (x^{\frac{5}{5}}-1)(x^{\frac{5}{5}}\zeta_5^5-1)...(x^{\frac{5}{5}}\zeta_5^20-1) \\ &= (x-1)^5 \textit{ in } \mathbb{F}_2[x]. \end{aligned}$$

Which means that x^5-1 factors in $\mathbb{F}_{2^5}[x,\theta]$ into linear terms. Sage [14] can confirm that:

$$x^5+1=(x+a^3)(x+a^4+a^3+1)(x+a^3+a+1)(x+a^4+a^2+a)(x+a^4+1)$$

for example. But this is not a unique factorisation. Any other factorisation will, however, have 5 linear terms by theorem 10.

When $n=9$, for example, we have in $\mathbb{F}_{2^2}[x,\theta]$

$$\begin{aligned} x^9-1 &= \prod_{i=0}^{1}[(x^{\frac{1}{2}}\zeta_2^i)^9-1] \\ &= (x^{\frac{9}{2}}-1)^2 \\ &= x^9-1 \textit{ in } \mathbb{F}_2[x] \\ &= (x+1)(x^2+x+1)(x^6+x^3+1). \end{aligned}$$

In $\mathbb{F}_{2^3}[x,\theta]$ we should have

$$\begin{aligned} x^9-1 &= \prod_{i=0}^{2}[(x^{\frac{1}{3}}\zeta_3^i)^9-1] \\ &= (x^{\frac{9}{3}}-1)(x^{\frac{9}{3}}\zeta_3^9-1)(x^{\frac{9}{3}}\zeta_3^{18}-1) \\ &= (x^3-1)^3 \textit{ in } \mathbb{F}_2[x] \\ &= (x+1)^3(x^2+x+1)^3. \end{aligned}$$

And indeed, this is one of factorisations of x^9-1 *in* $\mathbb{F}_{2^3}[x,\theta]$*:*

$$x^9+1=(x^2+(a+1)x+a^2+a+1)(x^2+a^2x+1)(x^2+x+a^2)(x+a)(x+1)(x+a^2+1).$$

Clearly, x^9-1 *will decompose into linear terms in* $\mathbb{F}_{2^9}[x,\theta]$*,* $\mathbb{F}_{2^{27}}[x,\theta]$ *etc. while in all other finite fields it will decompose uniquely as* x^9-1 *does in* $\mathbb{F}_2[x]$*.*

4.4 Jordan-Hölder and Factorisation in the Center

We state without a proof the following useful result of Jacobson [8].

Theorem 10. *If* $P\in\mathbb{F}_q[x,\theta]$ *has two decompositions into irreducible factors*

$$\begin{aligned}P&=P_1P_2...P_n\\&=\tilde{P}_1\tilde{P}_2...\tilde{P}_m\end{aligned}$$

then $n=m$ *and there exists a permutation* σ *such that* $P_i=\tilde{P}_{\sigma(i)}$ *with* $deg(P_i)=deg(\tilde{P}_{\sigma(i)})$*.*

This theorem of Jacobson is in fact more general then stated here but this version will suffice for our purposes.

Let $V\subset\overline{\mathbb{F}}_p$. Proposition 5 of section 4.1 established correspondence between θ^m-stable $\mathbb{F}_p$-vector space V and L_V, an element of $\mathbb{F}_q[\theta]$. Proposition 7 implied that V is irreducible under θ^m if and only if L_V is irreducible. Consider the following construction. Given V determine θ^m-stable subspace $V_1\neq V$ of maximal dimension.

Proposition 10. V/V_1 *is irreducible under* θ^m*.*

Proof. To see this suppose there exists θ^m-stable subspace $U\subset V/V_1$. Define a map

$$\pi:V\to V/V_1.$$

Then $\pi^{-1}(U)$ is θ^m-stable and contains V_1. The only choices are $\pi^{-1}(U)=V_1$ and $\pi^{-1}(U)=V$. If $\pi^{-1}(U)=V_1$ then $U=\pi(V_1)=\{0\}$ and if $\pi^{-1}(U)=V$ then $U=\pi(V)=V/V_1$. Thus V/V_1 is irreducible under θ^m. □

We can repeat the same procedure on V_1. If we continue doing that until we find an irreducible vector space we get the decomposition series

$$V=V_0\supset V_1\supset...\supset V_{r-1}\supset V_r=\{0\}\qquad(*)$$

that consists of θ^m-stable vector spaces V_i such that V_i/V_{i+1} is irreducible. This leads to a number of interesting remarks.

First of all, note that by Jordan-Hölder theorem for modules [9] any other decomposition series

$$V=V_0\supset V_1'\supset...\supset V_{r'-1}'\supset V_r'=\{0\}$$

is equivalent to $(*)$ in sense that $r=r'$ and there exists a permutation σ such that

$$V_i/V_{i+1}=V'_{\sigma(i)}/V'_{\sigma(i+1)}$$

for all i.

Proposition 11. *Theorem 10 follows from Jordan-Hölder.*

Proof. To see this we first show the correspondence between irreducible quotients V_i/V_{i+1} and irreducible factors P_i of polynomial P where $\mathrm{Sol}(P) = V$.

By proposition 5 there exist irreducible polynomials $P_r, P_{r-1}, ..., P_0$ such that

$$V_i = \mathrm{Sol}(P_i...P_r).$$

Define a map

$$\phi : \mathrm{Sol}(P_iP_{i+1}...P_n) \to \mathrm{Sol}(P_i)$$

that sends v in $\mathrm{Sol}(P_i...P_n)$ to $P_{i+1}...P_n(v)$ in $\mathrm{Sol}(P_i)$. Now, ϕ is surjective and $P_{i+1}...P_n(v)$ is clearly in $\mathrm{Sol}(P_i)$ since v being a root of $P_i...P_n$ implies that $P_{i+1}...P_n(v)$ is a root of P_i. Kernel of this map is $\mathrm{Sol}(P_{i+1}...P_n)$. By the first isomorphism theorem

$$\mathrm{Sol}(P_i...P_n)/\mathrm{Sol}(P_{i+1}...P_n) \simeq \mathrm{Sol}(P_i). \qquad (**)$$

Note that $dim(\mathrm{Sol}(P_i)) = deg(P_i)$. From section 4.1 remark 3 we know that $\mathrm{Sol}(P_i)$ is irreducible under the action of θ^m because P_i is irreducible for every i. Finally, by Jordan-Hölder any two decomposition series

$$\mathrm{Sol}(P) \supset \mathrm{Sol}(P_2...P_n) \supset ... \supset \mathrm{Sol}(P_{n-1}P_n) \supset \mathrm{Sol}(P_n) \supset \{0\}$$
$$\mathrm{Sol}(\tilde{P}) \supset \mathrm{Sol}(\tilde{P}_2...\tilde{P}_m) \supset ... \supset \mathrm{Sol}(\tilde{P}_{m-1}\tilde{P}_m) \supset \mathrm{Sol}(\tilde{P}_m) \supset \{0\}$$

are equivalent and there exists a permutation τ such that

$$\mathrm{Sol}(P_i) = \mathrm{Sol}(\tilde{P}_{\tau(i)}).$$

By proposition 5 of section 4.1 we are done. □

We close with the following theorem. Together with propositions of section 4.1 it solves the problem of predicting the degrees of irreducible factors of polynomials in $\mathbb{F}_q[x, \theta]$. It implies that any $L \in \mathbb{F}_q[\theta]$ decomposes into irreducible factors of the same degree as degrees of irreducible factors of $f_L \in \mathbb{F}_p[x]$.

Theorem 11. *Let $f_V(x)$ be a characteristic polynomial of θ^m on V. Then*

(i) for $W \subset V$, W θ^m-stable, we have $f_V = f_{V/W} f_W$

(ii) $f_V(x)$ is irreducible if and only if V is irreducible.

Proof. Let $v_1, ..., v_m$ denote a basis of W. We can complete this to a basis $v_1, ..., v_n$ of V. Since Both V and W are stable under θ^m we can write a matrix, denote it by M, of θ^m with respect to $v_1, ..., v_n$:

$$M = \begin{bmatrix} M_{\theta^m,W} & N \\ 0 & M_{\theta^m,V/W} \end{bmatrix}$$

where $M_{\theta^m,W}$ is an $m \times m$ matrix and $M_{\theta^m,V/W}$ is $(n-m) \times (n-m)$ matrix. The reason we have all zeros below $M_{\theta^m,W}$ is that W is θ^m-invariant and thus we want

$$Tv = a_1v_1 + ... + a_mv_m + a_{m+1}v_{m+1} + ... + a_nv_n$$

to be in W for all $v \in W$ and for that we must have $a_{m+1}, ..., a_n$ all zero. Now, to express the characteristic polynomial of M in term of characteristic polynomials of $M_{\theta^m,W}$ and $M_{\theta^m,V/W}$ note that we can rewrite M in the following way:

$$M = \begin{bmatrix} M_{\theta^m,W} & N \\ 0 & M_{\theta^m,V/W} \end{bmatrix} = \begin{bmatrix} M_{\theta^m,W} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & M_{\theta^m,W}^{-1} N \\ 0 & M_{\theta^m,V/W} \end{bmatrix}$$

to finally get

$$f_V = det(M) = det(M_{\theta^m,W}) det(M_{\theta^m,V/W}) = f_W f_{V/W}.$$

To prove part (ii) note that we have already seen one direction of this equivalence. Namely, in proposition 9 we saw that if $W \subset V$ and W is θ^m-stable then f_W divides f_V. The only thing left to show is that if f_V is reducible then V is reducible. So let g be a proper divisor of f_V with $deg(g) = d, deg(f_V) = n$ and $0 < d < n$. If $0 \subsetneq g(\theta^m)V \subsetneq V$ then we are done. Otherwise, $g(\theta^m)V$ is either equal to 0 or V. If $g(\theta^m)V = 0$ then let v be a vector in V and define $W = span(v, \theta^m v, ..., \theta^{m(d-1)} v)$. Then W is θ^m-stable and proper subspace of V, thus V is reducible. If on the other hand $g(\theta^m)V = V$ then

$$0 = f_v(\theta^m)V = h(\theta^m)g(\theta^m)V = h(\theta^m)V$$

where the first equality Holds because of Cayley-Hamilton and second one from f_V being reducible. Now we have that $h(\theta^m)V = 0$ and $deg(h) < n$ so we can repeat the same argument for h as we had for g. □

Corollary 1. *Let $f_V(x)$ be a characteristic polynomial of θ^m on V and*

$$V = V_0 \supset V_1 \supset ... \supset V_{r-1} \supset V_r = \{0\}$$

a decomposition series of V into θ^m-stable subspaces so that V_i/V_{i+1} is irreducible. Then

$$f_V(x) = \prod_{i=1}^{r-1} f_{V_i/V_{i+1}}(x)$$

and $f_{V_i/V_{i+1}}(x)$ irreducible.

Conclusion

Motivation for defining codes over skew polynomial rings was (in part) based on the fact that skew polynomial ring has a different structure than the commutative polynomial ring. Consequently, polynomials have very different factorisations. This factorisation is furthermore not unique. The larger the field over which we take polynomials, the larger the number of right divisors and thus the larger the number of generators of cyclic codes. Since there are many more generators in skew rings it makes sense to hope that we will come across abundance of simple and practical cyclic codes that improve the properties of already well known cyclic codes over commutative rings. As we have seen, however, polynomial $x^n - 1$ whose divisors generate cyclic codes, has a surprisingly simple decomposition over $\mathbb{F}_q[x, \theta]$. In fact it decomposes most of the time uniquely and in the same way as it does in $\mathbb{F}_p[x]$. So there are simply no new codes even for arbitrarily large m, the order of the automorphism θ. When $(m, n) \neq 1$, however, we do get non unique factorisation. Only in this case it makes sense to look for new better codes.

We finished with predicting the degrees of irreducible factors of all polynomials in $\mathbb{F}_q[x, \theta]$. A natural next step for research is to find the actual irreducible factors.

Bibliography

[1] F. Beukers, *Factorisation in skew rings*, Inofficial notes (2010).

[2] D. Boucher, W. Geiselmann, and F. Ulmer, *Skew-cyclic codes*, Appl. Algebra Eng., Commun. Comput. **18** (2007), 379–389.

[3] D. Boucher and F. Ulmer, *Coding with skew polynomial rings*, J. Symb. Comput. **44** (2009), 1644–1656.

[4] D.A. Cox, *Galois theory*, Pure and applied mathematics, Wiley-Interscience, 2004.

[5] E. M. Gabidulin, *Theory of Codes with Maximum Rank Distance*, Probl. Peredachi Inf. **21** (1985), no. 1, 3–16.

[6] M. Giesbrecht and Y. Zhang, *Factoring and decomposing ore polynomials over fq(t)*, Proceedings of the 2003 international symposium on Symbolic and algebraic computation (New York, NY, USA), ISSAC '03, ACM, 2003, pp. 127–134.

[7] R. W. Hamming, *Error detecting and error correcting codes*, Bell System Technical Journal, Vol. 26, No. 2 (1950), 147–160.

[8] N. Jacobson, *The theory of rings*, Mathematical surveys, American Mathematical Society, 1943.

[9] S. Lang, *Algebra*, Graduate texts in mathematics, Springer, 2002.

[10] Arjen K. Lenstra, *Factorization of polynomials*, SIGSAM Bull. **18** (1984), 16–18.

[11] R. Lidl and H. Niederreiter, *Introduction to finite fields and their applications*, Cambridge University Press, 1994.

[12] J.H. Lint, *Introduction to coding theory*, Graduate texts in mathematics, Springer, 1999.

[13] R.J. McEliece, *The theory of information and coding*, Cambridge University Press, 2002.

[14] W. Stein, *Sage, mathematics software*.