Algebraic geometry codes

Tom Høholdt, Jacobus H. van Lint and Ruud Pellikaan [*]

In the *Handbook of Coding Theory*, vol 1, pp. 871-961,
(V.S. Pless, W.C. Huffman and R.A. Brualdi Eds.),
Elsevier, Amsterdam 1998.

[*] The first author is from the Department of Mathematics, Technical University of Denmark, Bldg 303, DK 2800, Lyngby, Denmark. The last two authors are from the Department of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

Contents

1 Introduction

Consider a geometric object $\mathcal{X}$ with a subset $\mathcal{P}$ consisting of n points which are enumerated $P_1, \ldots, P_n$. Suppose that we have a vector space L over $\mathbb{F}_q$ of functions on $\mathcal{X}$ with values in $\mathbb{F}_q$. Thus $f(P_i) \in \mathbb{F}_q$ for all i and $f \in L$. In this way one has an evaluation map

$$ev_{\mathcal{P}} : L \longrightarrow \mathbb{F}_q^n$$

which is defined by $ev_{\mathcal{P}}(f) = (f(P_1), \ldots, f(P_n))$. This evaluation map is linear, so its image is a linear code. The image and its dual are the objects of study of this chapter. The dimension and the minimum distance of these codes and their duals will be considered. Decoding algorithms for these codes will be treated.

Defined in this generality, not much can be specifically said about the parameters of these codes. In the following, $\mathcal{X}$ is a subset of the affine or projective space which is the common set of zeros of some given set of polynomials, called a *variety* . $P_1, \ldots, P_n$ will be *rational points* of $\mathcal{X}$, i.e. points that have coordinates in $\mathbb{F}_q$. The functions will be polynomials or rational functions, that is to say quotients of polynomials. We call the above codes *algebraic geometry (AG)* codes if some theory of the variety $\mathcal{X}$ gives bounds on the dimension of the vector space L and the minimum distance of the code.

The classical example of the above situation is given by *Reed-Solomon (RS)* codes. Here the geometric object $\mathcal{X}$ is the affine line over $\mathbb{F}_q$, the points are n distinct elements of $\mathbb{F}_q$ and L is the vector space of polynomials of degree at most $k-1$ and with coefficients in $\mathbb{F}_q$. This vector space has dimension k. Such polynomials have at most $k-1$ zeros, so nonzero codewords have at least $n-k+1$ nonzeros. Hence this code has parameters $[n, k, n-k+1]$ if $k \leq n$. The length of a RS code is at most q. A way to get longer codes is by considering subfield subcodes or trace codes of RS codes. In this way one gets cyclic codes.

If we take as geometric object $\mathcal{X}$ the affine space of dimension m over $\mathbb{F}_q$, for the set $\mathcal{P}$ all the q^m points of this affine space, and as vector space all polynomials of degree at most r, then we get the *Reed-Muller (RM)* codes of order r in m variables over $\mathbb{F}_q$.

Every variety has a *dimension* and a variety of dimension one is called an *algebraic curve.* If $\mathcal{X}$ is an algebraic curve over $\mathbb{F}_q$, $\mathcal{P}$ a set of n distinct points of $\mathcal{X}$ that are defined over $\mathbb{F}_q$, and L a vector space of rational functions with prescribed behavior of their poles and zeros, then we get the *geometric Goppa* codes. The parameters of these codes are determined by the theorem of *Riemann-Roch* , and they satisfy the following bound

$$k + d \geq n + 1 - g, \quad \text{or equivalently} \quad d \geq n + 1 - k - g,$$

where g is an invariant of the curve called its *genus* . The best codes are obtained for curves of genus zero. They are in fact extended generalized RS codes. These codes have length at most $q + 1$ and are therefore not capable of giving asymptotically good sequences of codes. The length n of RM codes is not bounded, but k/n or d/n tends to zero if $n \to \infty$. The *information rate* $R = k/n$ and the $\delta = d/n$ of geometric Goppa codes satisfy the following inequality

$$R + \delta \geq 1 - \frac{g-1}{n}.$$

For good geometric Goppa codes, curves of low genus with many rational points are therefore needed. By studying the number of rational points on *modular curves* over finite fields it was shown that there exist asymptotically good sequences of geometric Goppa codes satisfying the *Tsfasman-Vlăduţ-Zink* (TVZ) bound

$$R + \delta \geq 1 - \frac{1}{\sqrt{q} - 1} \qquad \text{when } q \text{ is a square.}$$

This bound is better than the *Gilbert-Varshamov (GV)* bound when $q \geq 49$. It was the first time that the GV bound could be improved.

At the end of the eighties, an active period of research on decoding algorithms for AG codes started, when the decoding algorithm for RS codes was generalized. RS codes are decoded up to half their minimum distance by first finding the *error positions* as zeros of a polynomial known as the *error-locator* polynomial. If the error positions are known and their number is strictly smaller than the minimum distance, then error values can be obtained by solving linear equations involving *syndromes* . This idea was generalized by error-locator functions on curves. The resulting *basic algorithm* decodes up to half the designed minimum distance minus the genus. A technique called *majority voting of unknown syndromes* gives an algorithm which decodes up to half the designed minimum distance. Yet faster decoding algorithms were devised with an application of *linear recurring sequences* in several variables. This is a multivariate generalization of the algorithm of *Berlekamp-Massey* .

The theory of algebraic geometry codes is rather involved and deep. To treat algebraic curves (or equivalently *algebraic function fields* of one variable) in a self-contained way, is already beyond the scope of this chapter. A large part of the theory of *modular curves* is required to understand the result on the asymptotically good sequences of codes on these curves. The complexity of the construction of these codes is polynomial but, because the polynomial degree is high, still not suitable for practical applications.

Several attempts have been made to give an elementary treatment. That means an easier way both to construct the codes and to understand and prove their properties. For plane curves, the theorem of *Bézout* was used to compute the parameters of the codes, but for the dual codes the theorem of *Riemann-Roch* was still needed. The *majority voting* for *unknown syndromes* gave a new bound for the minimum distance. It was the starting point of an elementary treatment of AG codes and it is the foundation of the main part of this chapter.

It resulted furthermore in an explicit and easy description of asymptotically good sequences of curves over $\mathbb{F}_q$ when q is a square. Thus the theory has been simplified drastically, but it still needs the theory of *Artin-Schreier* extensions. The corresponding codes are not yet known by an explicit description, but a start has been made.

Our aim with this chapter is not to survey the vast body of literature on AG codes but to give an account of the construction and decoding of these codes which can be treated in a self-contained and elementary way.

The key concept in our treatment is the notion of an *order function* . This concept is well-known in the context of *computational algebra* and *Gröbner bases* where *reduction orders* on monomials are intensively used. Two more applications of order functions will be given: bounds on the minimum distance and decoding.

In Sections 3-7 the theory is developed for the class of *evaluation* codes and their duals, giving all the necessary definitions, theorems and proofs using only linear algebra and some elementary knowledge of the ring of polynomials in several variables as a background.

The class of evaluation codes with their duals contains codes on varieties of arbitrary dimension and therefore intersects the class of geometric Goppa codes in the set of so-called one point codes on curves.

The part on asymptotically good sequences of AG codes will only be outlined.

Section 2 contains an outline of the standard description of algebraic geometry codes. Section 3 introduces the concepts order and weight functions. Section 4 defines and proves bounds on the minimum distance of evaluation codes and their duals. Section 5 treats special order functions, which are called weight functions, and their associated semigroups. Properties on the minimum distance for the codes are shown. The decoding of AG codes is treated in Section 6, where the basic algorithm and the majority voting scheme of unknown syndromes is explained. Section 7 gives a fast decoding algorithm.

References are not included in the main text but each section ends with a subsection called Notes, where references and some history are given.

Notation: A field is denoted by $\mathbb{F}$ and its algebraic closure by $\bar{\mathbb{F}}$. The set of nonzero elements of $\mathbb{F}$ is denoted by $\mathbb{F}^*$. The field of real numbers is denoted by $\mathbb{R}$. The finite field with q elements is denoted by $\mathbb{F}_q$. The standard inner product on the vector space $\mathbb{F}^n$ is defined by $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^n x_i y_i$. The integers are denoted by $\mathbb{Z}$, the nonnegative integers by $\mathbb{N}_0$ and the positive integers by $\mathbb{N}$. The greatest common divisor of two integers a and b is denoted by $\gcd(a, b)$. An $\mathbb{F}$-algebra will be a commutative ring with a unit that has $\mathbb{F}$ as a unitary subring, and it will be denoted by R. Most of the time the $\mathbb{F}$-algebra will be $\mathbb{F}[X_1, \ldots, X_m]$, the polynomial ring in m variables with coefficients in $\mathbb{F}$, or its factor rings $\mathbb{F}[X_1, \ldots, X_m]/I$ where I is an ideal of $\mathbb{F}[X_1, \ldots, X_m]$. Elements of $\mathbb{F}[X_1, \ldots, X_m]$ will be denoted by capitals F, G and H and the corresponding cosets in $\mathbb{F}[X_1, \ldots, X_m]/I$ by f, g and h, respectively.

2 Codes from curves

Reed-Solomon codes can be defined by considering points with coordinates in $\mathbb{F}_q$ on the projective line. Codewords are defined by considering rational functions with a pole of restricted order at a specified point and taking the values of these functions at the given points as coordinates. The classical Goppa codes are defined by calculating residues of certain functions at given points. The set of functions is restricted by requirements on their zeros and poles. These two ideas are what we shall generalize in this section. We must study algebraic curves, find a way to describe the restrictions on the set of functions that we use, and generalize the concept of residue. We describe two classes of codes that are duals. Finally we consider asymptotically good codes on curves.

In this section the theory is outlined and most of the proofs are omitted.

2.1 Algebraic curves

In the following, $\mathbb{F}$ is an algebraically closed field. In our applications, $\mathbb{F}$ will be the algebraic closure of $\mathbb{F}_q$. $\mathbb{A}^n$ will denote n-dimensional affine space with coordinates $x_1, x_2, \ldots, x_n$. Similarly, $\mathbb{P}^n$ will be n-dimensional projective space with homogeneous coordinates $x_0, x_1, \ldots, x_n$. First, we discuss the affine case. The situation for projective spaces is slightly more complicated.

In the space $\mathbb{A}^n$, the *algebraic* sets are the sets of zeros of ideals I of $\mathbb{F}[X_1, X_2, \ldots, X_n]$, that is to say

$$B = V(I) = \{(x_1, x_2, \ldots, x_n) \in \mathbb{A}^n \quad | \quad F(x_1, x_2, \ldots, x_n) = 0 \text{ for all } F \in \ I\}.$$

We always assume that I is *radical*, this means that $F \in I$ if $F^n \in I$ for some $n \in \mathbb{N}_0$, so I consists of *all* the polynomials that vanish on B, by Hilbert's Nullstellensatz. An algebraic set B is called *irreducible* if B cannot be written as the union of two proper algebraic subsets of B. An ideal I is called *prime* if $F \in I$ or $G \in I$ for all F, G such that $FG \in I$. The set $V(I)$ is irreducible if and only if I is a prime ideal.

Example 2.1 In the affine plane, consider the principal ideal generated by $X^2 - Y^2$. The corresponding algebraic set is the union of two lines with equations $Y = X$, respectively $Y = -X$. Each of these lines is an irreducible algebraic set in the plane $\mathbb{A}^2$.

All the curves in affine or projective space in this paragraph are required to be irreducible.

Definition 2.2 Consider a prime ideal I in the ring $\mathbb{F}[X_1, X_2, \ldots, X_n]$. The set $\mathcal{X}$ of zeros of I is called an *affine variety.*

Example 2.3 In 3-dimensional space, we consider the unit sphere, that is to say, the set with equation $X^2 + Y^2 + Z^2 = 1$. In our terminology, this is the affine variety consisting of the zeros of the ideal I, generated by the polynomial $X^2 + Y^2 + Z^2 - 1$. We are just using algebraic terminology to describe geometric objects that are defined by equations.

Two polynomials that differ by an element of I will have the same value in each point of $\mathcal{X}$. This is the reason for introducing the following ring.

Definition 2.4 The ring $\mathbb{F}[X_1, X_2, \ldots, X_n]/I$ is called the *coordinate ring* $\mathbb{F}[\mathcal{X}]$ of the variety $\mathcal{X}$.

We adopt the convention to use capital letters $X_1, \ldots, X_n$, Y and Z to denote variables. Polynomials are denoted by F, G and H and their cosets modulo the ideal I are denoted by small letters f, g and h, respectively.

The coordinate ring is an *integral domain*, that is to say, $f = 0$ or $g = 0$ for all f, g such that $fg = 0$, since I is a prime ideal. Therefore, we can make the following definition.

Definition 2.5 The quotient field of the ring $\mathbb{F}[\mathcal{X}]$ is denoted by $\mathbb{F}(\mathcal{X})$. It is called the *function field* of $\mathcal{X}$. The elements of $\mathbb{F}(\mathcal{X})$ are called *rational functions*. The *dimension* of the variety $\mathcal{X}$ is the transcendence degree of $\mathbb{F}(\mathcal{X})$ over $\mathbb{F}$. If this dimension is 1, $\mathcal{X}$ is called an *algebraic curve* .

Example 2.6 In the affine plane over the field $\mathbb{F}$, we consider the parabola $\mathcal{X}$ with equation $Y^2 = X$. In this example, the coordinate ring $\mathbb{F}[\mathcal{X}]$ consists of all the expressions of the form $A + By$, where A and B are in $\mathbb{F}[x]$ and y satisfies $y^2 = x$. So, $\mathbb{F}(\mathcal{X})$ is an algebraic extension of $\mathbb{F}(x)$ by the element y, satisfying this equation of degree 2.

In projective space $\mathbb{P}^n$, the situation is complicated by the fact that we must use homogeneous coordinates. A point $(x_0 : x_1 : \cdots : x_n)$ in $\mathbb{P}^n$ is the line in $\mathbb{A}^{n+1}$ through the origin and $(x_0, x_1, \ldots, x_n) \neq 0$. So $(x_0 : x_1 : \cdots : x_n) = (y_0 : y_1 : \cdots : y_n)$ if and only if $(x_0, x_1, \ldots, x_n) = \lambda(y_0, y_1, \ldots, y_n)$ for some $\lambda \in \mathbb{F}^*$. Hence it makes sense to consider the zero set in $\mathbb{P}^n$ of homogeneous polynomials, but for rational functions to have a meaning one takes only those quotients for which numerator and denominator are homogeneous polynomials of the same degree. A projective variety $\mathcal{X}$ is the zero set in $\mathbb{P}^n$ of a homogeneous prime ideal I in $\mathbb{F}[X_0, X_1, \ldots, X_n]$. Consider the subring $R(\mathcal{X})$ of $\mathbb{F}(X_0, X_1, \ldots, X_n)$ consisting of the fractions F/G, where F and G are homogeneous polynomials of the same degree and $G \notin I$. Then $R(\mathcal{X})$ has a unique maximal ideal $M(\mathcal{X})$ consisting of all those F/G with $F \in I$. The function field $\mathbb{F}(\mathcal{X})$ is by definition $R(\mathcal{X})/M(\mathcal{X})$.

Definition 2.7 Let $\mathcal{X}$ be an affine or a projective variety. Let P be a point on $\mathcal{X}$. Then a rational function ϕ is called *regular* in the point P if one can find polynomials F and G, respectively homogeneous polynomials of the same degree, such that $G(P) \neq 0$ and ϕ is the coset of F/G. The functions that are regular in every point of the set U form a ring, denoted by $\mathbb{F}[U]$.

If $\mathcal{X}$ is affine, then the coordinate ring of $\mathcal{X}$ coincides with the ring of regular functions on $\mathcal{X}$, so there is no ambiguity in the notation $\mathbb{F}[\mathcal{X}]$.

If $\mathcal{X}$ is projective, then there are no regular functions on $\mathcal{X}$ except constant functions.

Definition 2.8 The *local ring* $\mathcal{O}_P$ (sometimes denoted by $\mathcal{O}_P(\mathcal{X})$) of the point P on the variety $\mathcal{X}$ is the set of rational functions that are regular in P.

The reader familiar with algebraic terminology will realize that this is indeed a "local ring" in the algebraic sense, that is to say,it has a unique maximal ideal, namely the set $\mathcal{M}_P$ of functions in $\mathcal{O}_P$ that are zero in P.

Example 2.9 In $\mathbb{P}^2$ with coordinates $(x : y : z)$, consider the variety $\mathcal{X}$ defined by $XZ - Y^2 = 0$. This is the parabola of Example 2.6, now with one point at infinity, namely Q=(1:0:0). The function x/y is equal to y/z on the curve, hence it is regular in the point $P = (0 : 0 : 1)$. The function $(2xz + z^2)/(y^2 + z^2)$ is regular in P. By replacing y^2 by xz, we see that the function is equal to $(2x + z)/(x + z)$ and therefore also regular in Q.

Now we show how to embed an affine variety in a projective variety. Associate with $F \in \mathbb{F}[X_1, \ldots, X_n]$ the homogeneous polynomial F^* defined by

$$F^* = X_0^l F(X_1/X_0, \ldots, X_n/X_0),$$

where l is the degree of F. Let $\mathcal{X}$ be an affine variety in $\mathbb{A}^n$ defined by the prime ideal I. Let I^* be the ideal generated by $\{F^* | F \in I\}$. Then I^* is a homogeneous prime ideal defining the projective variety $\mathcal{X}^*$ in $\mathbb{P}^n$. Let $\mathcal{X}_0^* = \{(x_0 : x_1 : \cdots : x_n) \in \mathcal{X}^* | x_0 \neq 0\}$. Then $\mathcal{X}$ is isomorphic with $\mathcal{X}_0^*$ under the map $(x_1, \ldots, x_n) \mapsto (1 : x_1 : \cdots : x_n)$. The points $(x_0 : x_1 : \cdots : x_n) \in \mathcal{X}^*$ such that $x_0 = 0$ are called *the points at infinity* of $\mathcal{X}$. Furthermore the function fields $\mathbb{F}(\mathcal{X})$ and $\mathbb{F}(\mathcal{X}^*)$ are isomorphic under the map $f/g \mapsto f^* x_0^m / g^*$, where $m = \deg(g) - \deg(f)$.

Conversely, for any point P of a projective variety $\mathcal{X}$ and any hyperplane $\mathcal{H}$ not containing P the complement $\mathcal{X} \setminus \mathcal{H}$ is an affine variety containing P.

From now on, most of the time we will consider plane curves to simplify the treatment.

Definition 2.10 Let $F = \sum a_{ij} X^i Y^j \in \mathbb{F}[X, Y]$. Then F_X, the *partial derivative* of F *with respect to* X is defined by

$$F_X = \sum i a_{ij} X^{i-1} Y^j$$

and F_Y is defined similarly.

Definition 2.11 Consider a curve $\mathcal{X}$ in $\mathbb{A}^2$, defined by the equation $F = 0$. Let P be a point on this curve. If at least one of the derivatives F_X or F_Y is not zero in P, then P is called a *simple* or *nonsingular* point of the curve. A curve is called *nonsingular*, *regular* or *smooth* if all the points are nonsingular.

Let $P = (a, b)$ be a nonsingular point on $\mathcal{X}$. The *tangent line* T_P at P is defined by $d_P F = 0$, where we define

$$d_P F = F_X(a, b)(X - a) + F_Y(a, b)(Y - b).$$

The definitions for a projective plane curve are similar and as follows. Let a projective plane curve be defined by the homogeneous equation $F = 0$. Let P be a point on this curve. If at least one of the derivatives F_X, F_Y or F_Z is not zero in P, then P is called a simple or nonsingular point of the curve.

Let $P = (a : b : c)$ be a nonsingular point of the curve. Then the tangent line at P has equation

$$F_X(a, b, c)X + F_Y(a, b, c)Y + F_Z(a, b, c)Z = 0.$$

Example 2.12 The *Fermat curve* $\mathcal{F}_m$ is a projective plane curve with defining equation

$$X^m + Y^m + Z^m = 0.$$

The partial derivatives of $X^m + Y^m + Z^m$ are mX^{m-1}, mY^{m-1}, and mZ^{m-1}. So considered as a curve over the finite field $\mathbb{F}_q$, it is regular if m is relatively prime to q.

Example 2.13 Let $q = r^2$. The *Hermitian curve* $\mathcal{H}_r$ over $\mathbb{F}_q$ is defined by the affine equation

$$U^{r+1} + V^{r+1} + 1 = 0.$$

The corresponding homogeneous equation is

$$U^{r+1} + V^{r+1} + W^{r+1} = 0.$$

Hence it has $r + 1$ points at infinity and it is the Fermat curve $\mathcal{F}_m$ over $\mathbb{F}_q$ with $r = m - 1$. The conjugate of $a \in \mathbb{F}_q$ over $\mathbb{F}_r$ is obtained by $\bar{a} = a^r$. So the equation can also be written as

$$U\bar{U} + V\bar{V} + W\bar{W} = 0.$$

This looks like equating a Hermitian form over the complex numbers to zero and explains the terminology.

We will see in Section 3 that for certain constructions of codes on curves it is convenient to have exactly one point at infinity. We will give a transformation such that the new equation of the Hermitian curve has this property. Choose an element $b \in \mathbb{F}_q$ such that $b^{r+1} = -1$. There are exactly $r + 1$ of these, since $q = r^2$. Let $P = (1 : b : 0)$. Then P is a point of the Hermitian curve. The tangent line at P has equation $U + b^r V = 0$. Multiplying with b gives the equation $V = bU$. Substituting $V = bU$ in the defining equation of the curve gives that $W^{r+1} = 0$. So P is the only intersection point of the Hermitian curve and the tangent line at P. New homogeneous coordinates are chosen such that this tangent line becomes the line at infinity. Let $X_1 = W$, $Y_1 = U$ and $Z_1 = bU - V$. Then the curve has homogeneous equation

$$X_1^{r+1} = b^r Y_1^r Z_1 + bY_1 Z_1^r - Z_1^{r+1}$$

in the coordinates X_1, Y_1 and Z_1. Choose an element $a \in \mathbb{F}_q$ such that $a^r + a = -1$. There are r of these. Let $X = X_1$, $Y = bY_1 + aZ_1$ and $Z = Z_1$. Then the curve has homogeneous equation

$$X^{r+1} = Y^r Z + YZ^r$$

with respect to X, Y and Z. Hence the Hermitian curve has affine equation

$$X^{r+1} = Y^r + Y$$

with respect to X and Y. This last equation has $(0 : 1 : 0)$ as the only point at infinity.

Example 2.14 The *Klein curve* has homogeneous equation

$$X^3Y + Y^3Z + Z^3X = 0.$$

More generally we define the curve $\mathcal{K}_m$ by the equation

$$X^mY + Y^mZ + Z^mX = 0.$$

Suppose that $m^2 - m + 1$ is relatively prime to q. The partial derivatives of the left side of the equation are $mX^{m-1}Y + Z^m$, $mY^{m-1}Z + X^m$ and $mZ^{m-1}X + Y^m$. Let $(x : y : z)$ be a singular point of the curve $\mathcal{K}_m$. If m is divisible by the characteristic, then $x^m = y^m = z^m = 0$. So $x = y = z = 0$, a contradiction. If m and q are relatively prime, then $x^my = -my^mz = m^2z^mx$. So

$$(m^2 - m + 1)z^mx = x^my + y^mz + z^mx = 0.$$

Therefore $z = 0$ or $x = 0$, since $m^2 - m + 1$ is relatively prime to the characteristic. But $z = 0$ implies $x^m = -my^{m-1}z = 0$. Furthermore $y^m = -mz^{m-1}x$. So $x = y = z = 0$, which is a contradiction. Similarly $x = 0$ leads to a contradiction. Hence $\mathcal{K}_m$ is nonsingular if $gcd(m^2 - m + 1, q) = 1$.

2.2 Local parameters and discrete valuations

We want to show that the maximal ideal $\mathcal{M}_P$ is a principal ideal, that is to say, generated by one element. Let $\mathcal{X}$ be a smooth curve in $\mathbb{A}^2$ defined by the equation $F = 0$, and let $P = (a, b)$ be a point on $\mathcal{X}$. The maximal ideal $\mathcal{M}_P$ is generated by $x - a$ and $y - b$. Now

$$F_X(P)(x-a) + F_Y(P)(y-b) \;\equiv\; 0 \bmod \mathcal{M}_P^2.$$

Hence the $\mathbb{F}$-vector space $\mathcal{M}_P/\mathcal{M}_P^2$ has dimension 1 and therefore $\mathcal{M}_P$ has one generator. Let $g \in \mathbb{F}[\mathcal{X}]$ be the coset of a polynomial G. Then g is a generator of $\mathcal{M}_P$ if and only if d_PG is not a constant multiple of d_PF.

Definition 2.15 Let t be a generating element of $\mathcal{M}_P$. We can then write every element z of $\mathcal{O}_P$ in a unique way as $z = ut^m$, where u is a unit and $m \in \mathbb{N}_0$. The function t is called a *local parameter* or *uniformizing parameter* in P. If $m > 0$, then P is a *zero* of multiplicity m of z. We write $m = \mathrm{ord}_P(z) = v_P(z)$. We use the convention $v_P(0) = \infty$.

Theorem 2.16 *The map $v_P : \mathcal{O}_P \to \mathbb{N}_0 \cup \{\infty\}$ is a discrete valuation, that is to say the map is surjective and it satisfies the following properties:*

(i) $v_p(f) = \infty$ *if and only if* $f = 0$
(ii) $v_P(\lambda f) = v_P(f)$ *for all nonzero* $\lambda \in \mathbb{F}$
(iii) $v_P(f+g) \geq min\{v_P(f), v_P(g)\}$
and equality holds when $v_P(f) \neq v_P(g)$.
(iv) $v_P(fg) = v_P(f) + v_P(g)$
(v) *If* $v_P(f) = v_P(g)$, *then there exists a nonzero* $\lambda \in \mathbb{F}$ *such that* $v_P(f - \lambda g) > v_P(g)$.

for all $f, g \in \mathcal{O}_P$. Here $\infty > n$ for all $n \in \mathbb{N}_0$.

We extend the function v_P to $\mathbb{F}(\mathcal{X})$ by defining $v_P(f/g) = v_P(f) - v_P(g)$. If $v_P(z) = -m < 0$, then we say that z has a *pole* of order m in P. If z is an element of $\mathbb{F}(\mathcal{X})$ with $v_P(z) = m$, then we can write $z = at^m + z'$, where $a \in \mathbb{F}$, $a \neq 0$, and $v_P(z') > m$. In this way, one can show that z can be expanded as a Laurent series $\sum_{i \geq m} a_i t^i$, where $a_i \in \mathbb{F}$ for all i and $a_m \neq 0$.

Example 2.17 Let $\mathbb{P}^1$ be the projective line over $\mathbb{F}$. A local parameter in the point $P = (1:0)$ is y/x. The rational function $(x^2 - y^2)/y^2$ has a pole of order 2 in P. If $\mathbb{F}$ does not have characteristic 2, then $(1:1)$ and $(-1:1)$ are zeros with multiplicity 1.

Example 2.18 Let the characteristic of $\mathbb{F}$ be unequal to 2. Let $\mathcal{X}$ be the circle in $\mathbb{A}^2$ with equation $X^2 + Y^2 = 1$ and let $P = (1, 0)$. Let $z = 1 - x$. This function is 0 in P, so it is in $\mathcal{M}_P$. We claim that z has order 2. To see this, observe that y is a local parameter in P, because the line $d_P Y = Y = 0$ is not equal to the tangent line $X = 1$ in P. Furthermore, on $\mathcal{X}$ we have $1 - x = y^2/(1 + x)$ and the funcion $1/(1 + x)$ is a unit in $\mathcal{O}_P$.

When we construct codes, we will be interested in points that have their coordinates in our alphabet $\mathbb{F}_q$. We give these a special name.

Definition 2.19 Let $\mathcal{X}$ be a curve defined over $\mathbb{F}_q$, that is to say, the defining equations have coefficients in $\mathbb{F}_q$. Then points on $\mathcal{X}$ with all their coordinates in $\mathbb{F}_q$ are called *rational points.*

Example 2.20 Consider the Klein quartic with equation $X^3Y + Y^3Z + Z^3X = 0$ of Example 2.14 over the algebraic closure of $\mathbb{F}_2$. Look at a few of the subfields. Over $\mathbb{F}_2$ the rational points are (1:0:0), (0:1:0), and (0:0:1). If we go to $\mathbb{F}_4$, there are two more rational points, namely $(1 : \alpha : 1 + \alpha)$ and $(1 : 1 + \alpha : \alpha)$ if $\mathbb{F}_4 = \{0, 1, \alpha, \alpha^2\}$, where $\alpha^2 = 1 + \alpha$.

In later examples, this curve will be studied over $\mathbb{F}_8$. As usual, we define this field as $\mathbb{F}_2(\xi)$, where $\xi^3 = \xi + 1$. If a rational point has a coordinate 0, it must be one of the points over $\mathbb{F}_2$. If $xyz \neq 0$, we can take $z = 1$. If $y = \xi^i$ $(0 \leq i \leq 6)$, then write $x = \xi^{3i}\eta$. Substitution in the equation gives $\eta^3 + \eta + 1 = 0$, that is to say, η is one of the elements ξ, ξ^2, or ξ^4. So we find a total of 24 rational points over $\mathbb{F}_8$.

Example 2.21 Let $\mathcal{X}$ be the plane curve with equation $X^3 + Y^3 + Z^3 = 0$ over the closure of $\mathbb{F}_2$ and look at the subfield $\mathbb{F}_4$. So we consider the Hermitian curve $\mathcal{H}_2$ of Example 2.13. Since a third power of an element of $\mathbb{F}_4$ is 0 or 1, all the rational points have one coordinate 0. We can take one of the others to be 1, and the third one any nonzero element of $\mathbb{F}_4$. So we find nine (projective) points. In $Q = (0 : 1 : 1)$, we can take $t = x/z$ as local parameter. We consider a difficulty that will come up again. The expression $f = x/(y + z)$ looks like a perfectly reasonable function and in fact on most of $\mathcal{X}$ it is. However, in Q the fraction does not make sense. We must find an equivalent form for f near Q. On $\mathcal{X}$ we have

$$\frac{x}{y + z} = \frac{x(y^2 + yz + z^2)}{y^3 + z^3} = t^{-2} \cdot \frac{y^2 + yz + z^2}{z^2},$$

where the second factor on the right is regular and not 0 in Q. By our earlier conventions, we say that f has a pole of order 2 in Q. Similarly, $y/(y+z)$ has a pole of order 3 in Q.

2.3 Bézout's theorem

We now consider the intersection of a curve and a hypersurface in $\mathbb{P}^n$. We assume that the reader is familiar with the fact that a polynomial of degree m in one variable, with coefficients in a field has at most m zeros. If the field is algebraically closed and if the zeros are counted with multiplicities, then the number of zeros is equal to m. We shall now state a theorem, known as *Bézout's theorem*, which is a generalization of these facts to polynomials in several variables.

The *degree* of a projective curve is the maximal number of points in the intersection with a hyperplane not containing the curve. So the degree of a projective plane curve is equal to the degree of the defining equation.

We only consider the intersection of an irreducible nonsingular projective curve $\mathcal{X}$ of degree l and a hypersurface $\mathcal{Y}$ defined by the equation $G = 0$ of degree m. We assume that $\mathcal{X}$ is not contained in $\mathcal{Y}$.

Definition 2.22 Let P be a point of $\mathcal{X}$. Let H be a homogeneous linear form such that $H(P) \neq 0$. Let h be the class of H modulo the ideal defining $\mathcal{X}$. Then the *intersection multiplicity* $I(P; \mathcal{X}, \mathcal{Y})$ of $\mathcal{X}$ and $\mathcal{Y}$ in P is defined by $v_P(g/h^m)$.

This definition does not depend on the choice of H, since h/h' is a unit in $\mathcal{O}_P$ for any other choice of a linear form H' that is not zero in P.

Theorem 2.23 *Let $\mathcal{X}$ be an irreducible nonsingular projective curve of degree l and $\mathcal{Y}$ a hypersurface of degree m in $\mathbb{P}^n$ such that $\mathcal{X}$ is not contained in $\mathcal{Y}$. Then they intersect in exactly lm points (if counted with multiplicity).*

We do not prove this theorem. If $\mathbb{F}$ is not algebraically closed or the curves are affine, then the curves intersect in at most lm points.

We mention two consequences of this theorem.

Corollary 2.24 *Two projective plane curves of positive degree have a point in common.*

Corollary 2.25 *A regular projective plane curve is irreducible.*

Proof. If $F = GH$ is a factorization of F with factors of positive degree, we get

$$F_X = G_X H + G H_X$$

by the product or Leibniz rule for the partial derivative. So F_X is an element of the ideal generated by G and H, and similarly for the other two partial derivatives. Hence the set of common zeros of F_X, F_Y, F_Z and F contains the set of common zeros of G and H. The intersection of the curves with equations $G = 0$ and $H = 0$ is not empty since G and H have positive degrees, by Corollary 2.24. Therefore the curve has a singular point. □

Remark 2.26 Notice that the assumption that the curve is a projective plane curve is essential. The equation $X^2Y - X = 0$ defines a regular affine plane curve, but is clearly reducible. However one gets immediately from Corollary 2.25 that if $F = 0$ is an affine plane curve and the homogenization F^* defines a regular projective curve, then F is absolutely irreducible. The affine curve with equation $X^2Y - X = 0$ has the points $(1 : 0 : 0)$ and $(0 : 1 : 0)$ at infinity, and $(0 : 1 : 0)$ is a singular point.

Let V_l be the vector space of polynomials of degree at most l in two variables X, Y and coefficients in $\mathbb{F}_q$. Consider an element G of degree m in $\mathbb{F}_q[X, Y]$ such that the homogeneous form G^* defines a nonsingular curve. Then G is irreducible in $\mathbb{F}[X, Y]$, where $\mathbb{F}$ is the algebraic closure of $\mathbb{F}_q$ by Corollary 2.25. Let $P_1, P_2, \ldots, P_n$ be rational points on the plane curve defined by the equation $G = 0$, that is to say, $P_i = (a_i, b_i) \in \mathbb{F}_q^2$ and $G(P_i) = 0$ for $1 \leq i \leq n$. We define a code C by

$$C = \{(F(P_1), F(P_2), \ldots, F(P_n)) \mid F \in \mathbb{F}_q[X, Y],\ \deg(F) \leq l\}.$$

We shall use d for the minimum distance of this code and (as usual) call the dimension k.

Theorem 2.27 *Let $n > lm$. For the minimum distance d and the dimension k of C, we have*

$$d \geq n - lm,$$

$$k = \begin{cases} \binom{l+2}{2} & \text{if } l < m, \\ lm + 1 - \binom{m-1}{2} & \text{if } l \geq m. \end{cases}$$

Proof. The monomials of the form $X^\alpha Y^\beta$ with $\alpha + \beta \leq l$ form a basis of V_l. Hence V_l has dimension $\binom{l+2}{2}$.

Let $F \in V_l$. If G is a factor of F, then the codeword in C corresponding to F is zero. Conversely, if this codeword is zero, then the curves with equation $F = 0$ and $G = 0$ have degree $l' \leq l$ and m respectively, and they have the n points $P_1, P_2, \ldots, P_n$ in their intersection. Bézout's theorem and the assumption $lm < n$ imply that F and G have a common factor. Since G is irreducible, F must be divisible by G. Hence the functions $F \in V_l$ that yield the zero codeword form the subspace GV_{l-m}. This implies that if $l < m$, then $k = \binom{l+2}{2}$, and if $l \geq m$, then

$$k = \binom{l+2}{2} - \binom{l-m+2}{2} = lm + 1 - \binom{m-1}{2}.$$

The same argument with Bézout's theorem shows that a nonzero codeword has at most lm coordinates equal to 0, that is to say, it has weight at least $n - lm$. Hence $d \geq n - lm$.

□

Remark 2.28 If $F_1, \ldots, F_k$ is a basis for V_l modulo GV_{l-m}, then

$$(F_i(P_j) \mid 1 \leq i \leq k, 1 \leq j \leq n)$$

is a generator matrix of C. So it is a parity check matrix for the dual of C. The minimum distance $d^\perp$ of $C^\perp$ is equal to the minimal number of dependent columns of this matrix. Hence for all $t < d^\perp$ and every subset $\mathcal{Q}$ of $\mathcal{P} = \{P_1, \ldots, P_n\}$ consisting of t distinct points, the corresponding $k \times t$ submatrix has maximal rank t. Let $L_l = V_l/GV_{l-m}$. Then the map that evaluates polynomials at the points of $\mathcal{Q}$ induces a surjective map from L_l to $\mathbb{F}_q^t$. The kernel, which we denote by $L_l(\mathcal{Q})$, is the space of all functions $F \in V_l$ that are zero at the points of $\mathcal{Q}$ modulo GV_{l-m}, So $\dim(L_l(\mathcal{Q})) = k - t$ if $t < d^\perp$.

Conversely, the dimension of $L_l(\mathcal{Q})$ is at least $k - t$ for all t-subsets $\mathcal{Q}$ of $\mathcal{P}$. But in order to get a bound for $d^\perp$, we have to know that $\dim(L_l(\mathcal{Q})) = k - t$ for all $t < d^\perp$. The theory developed so far is not sufficient to get such a bound. The theorem of Riemann-Roch gives an answer to this question. See Section 2.7. Section 4 gives another, more elementary, solution to this problem.

Notice that the following inequality holds for the code C:

$$k + d \geq n + 1 - g,$$

where $g = (m-1)(m-2)/2$. In Section 2.4 we will see that g is the genus, see Definition 2.48. In Sections 3-6 the role of g will be played by the number of gaps of the Weierstrass semigroup of a point at infinity, see Definition 2.60.

2.4 Divisors

In the following, $\mathcal{X}$ is an irreducible smooth projective curve over an algebraically closed field $\mathbb{F}$.

Definition 2.29 A *divisor* is a formal sum $D = \sum_{P \in X} n_P P$, with $n_P \in \mathbb{Z}$ and $n_P = 0$ for all but a finite number of points P. The *support* of a divisor is the set of points with nonzero coefficient. A divisor D is called *effective* if all coefficients n_P are nonnegative (notation $D \succcurlyeq 0$). The *degree* $\deg(D)$ of the divisor D is $\sum n_P$.

Definition 2.30 Let $\mathcal{X}$ and $\mathcal{Y}$ be projective plane curves defined by the equations $F = 0$ and $G = 0$, respectively, then the *intersection divisor* $\mathcal{X} \cdot \mathcal{Y}$ is defined by

$$\mathcal{X} \cdot \mathcal{Y} = \sum I(P; \mathcal{X}, \mathcal{Y}) P,$$

where $I(P; \mathcal{X}, \mathcal{Y})$ is the intersection mulitplicity of Definition 2.22.

Bézout's theorem tells us that $\mathcal{X} \cdot \mathcal{Y}$ is indeed a divisor and that its degree is lm if the degrees of $\mathcal{X}$ and $\mathcal{Y}$ are l and m, respectively.
Let $v_P = \mathrm{ord}_P$ be the discrete valuation defined for functions on $\mathcal{X}$ in Definition 2.15.

Definition 2.31 If f is a rational function on $\mathcal{X}$, not identically 0, we define the divisor of f to be

$$(f) = \sum_{P \in X} v_P(f) P.$$

So, in a sense, the divisor of f is a bookkeeping device that tells us where the zeros and poles of f are and what their multiplicities and orders are.

Theorem 2.32 *The degree of a divisor of a rational function is 0.*

Proof. Let $\mathcal{X}$ be a projective curve of degree l. Let f be a rational function on the curve $\mathcal{X}$. Then f is represented by a quotient A/B of two homogeneous polynomials of the same degree, say m. Let $\mathcal{Y}$ and $\mathcal{Z}$ be the hypersurfaces defined by the equations $A = 0$ and $B = 0$, respectively. Then $v_P(f) = I(P; \mathcal{X}, \mathcal{Y}) - I(P; \mathcal{X}, \mathcal{Z})$, since $f = a/b = (a/h^m)(b/h^m)^{-1}$, where H is a homogeneous linear form representing h such that $H(P) \neq 0$. Hence

$$(f) = \mathcal{X} \cdot \mathcal{Y} - \mathcal{X} \cdot \mathcal{Z}.$$

So (f) is indeed a divisor and its degree is zero, since it is the difference of two intersection divisors of the same degree lm. □

Example 2.33 Look at the curve of Example 2.21. We saw that $f = x/(y+z)$ has a pole of order 2 in $Q = (0 : 1 : 1)$. The line $\mathcal{L}$ with equation $X = 0$ intersects the curve in three points, namely $P_1 = (0 : \alpha : 1)$, $P_2 = (0 : 1+\alpha : 1)$ and Q. So $\mathcal{X} \cdot \mathcal{L} = P_1 + P_2 + Q$. The line $\mathcal{M}$ with equation $Y = 0$ intersects the curve in three points, namely $P_3 = (1 : 0 : 1)$, $P_4 = (\alpha : 0 : 1)$ and $P_5 = (1 + \alpha : 0 : 1)$. So $\mathcal{X} \cdot \mathcal{M} = P_3 + P_4 + P_5$. The line $\mathcal{N}$ with equation $Y + Z = 0$ intersects the curve only in Q. So $\mathcal{X} \cdot \mathcal{N} = 3Q$. Hence $(x/(y + z)) = P_1 + P_2 - 2Q$ and $(y/(y + z)) = P_3 + P_4 + P_5 - 3Q$.

In this example it is not necessary to compute the intersection multiplicities since they are a consequence of Bézout's theorem.

Example 2.34 Let $\mathcal{X}$ be the Klein quartic with equation $X^3Y + Y^3Z + Z^3X = 0$ of Example 2.14. Let $P_1 = (0 : 0 : 1)$, $P_2 = (1 : 0 : 0)$ and $Q = (0 : 1 : 0)$. Let $\mathcal{L}$ be the line with equation $X = 0$. Then $\mathcal{L}$ intersects $\mathcal{X}$ in the points P_1 and Q. Since $\mathcal{L}$ is not tangent in Q, we see that $I(Q; \mathcal{X}, \mathcal{L}) = 1$. So the intersection multiplicity of $\mathcal{X}$ and $\mathcal{L}$ in P_1 is 3, since the multiplicities add up to 4. Hence $\mathcal{X} \cdot \mathcal{L} = 3P_1 + Q$. Similarly we get for the lines $\mathcal{M}$ and $\mathcal{N}$ with equations $Y = 0$ and $Z = 0$, respectively, $\mathcal{X} \cdot \mathcal{M} = 3P_2 + P_1$ and $\mathcal{X} \cdot \mathcal{N} = 3Q + P_2$. Therefore $(x/z) = 3P_1 - P_2 - 2Q$ and $(y/z) = P_1 + 2P_2 - 3Q$.

Definition 2.35 The divisor of a rational function is called a *principal divisor*. We shall call two divisors D and D' *linearly equivalent* if and only if $D - D'$ is a principal divisor ; notation $D \equiv D'$.

This is indeed an equivalence relation.

Definition 2.36 Let D be a divisor on a curve $\mathcal{X}$. We define a vector space $\mathcal{L}(D)$ over $\mathbb{F}$ by

$$\mathcal{L}(D) = \{f \in \mathbb{F}(\mathcal{X})^* \mid (f) + D \succcurlyeq 0\} \cup \{0\}.$$

The dimension of $\mathcal{L}(D)$ over $\mathbb{F}$ is denoted by $l(D)$.

Note that if $D = \sum_{i=1}^{r} n_i P_i - \sum_{j=1}^{s} m_j Q_j$ with all $n_i, m_j > 0$, then $\mathcal{L}(D)$ consists of 0 and the functions in the function field that have zeros of multiplicity at least m_j at Q_j $(1 \leq j \leq s)$ and that have no poles except possibly at the points P_i, with order at most n_i $(1 \leq i \leq r)$. We shall show that this vector space has finite dimension.

First we note that if $D \equiv D'$ and g is a rational function with $(g) = D - D'$, then the map $f \mapsto fg$ shows that $\mathcal{L}(D)$ and $\mathcal{L}(D')$ are isomorphic.

Theorem 2.37
(i) $l(D) = 0$ *if* $deg(D) < 0$,
(ii) $l(D) \leq 1 + deg(D)$.

Proof. (i) If $\deg(D) < 0$, then for any function $f \in \mathbb{F}(\mathcal{X})^*$, we have $\deg((f) + D) < 0$, that is to say, $f \notin \mathcal{L}(D)$.

(ii) If f is not 0 and $f \in \mathcal{L}(D)$, then $D' = D + (f)$ is an effective divisor for which $\mathcal{L}(D')$ has the same dimension as $\mathcal{L}(D)$ by our observation above. So without loss of generality D is effective, say $D = \sum_{i=1}^{r} n_i P_i$, $(n_i \geq 0$ for $1 \leq i \leq r)$. Again, assume that f is not 0 and $f \in \mathcal{L}(D)$. In the point P_i, we map f onto the corresponding element of the n_i-dimensional vector space $(t_i^{-n_i} \mathcal{O}_{P_i}) / \mathcal{O}_{P_i}$, where t_i is a local parameter at P_i. We thus obtain a mapping of f onto the direct sum of these vector spaces ; (map the 0-function onto 0). This is a linear mapping. Suppose that f is in the kernel. This means that f does not have a pole in any of the points P_i, that is to say, f is a constant function. It follows that

$$l(D) \leq 1 + \sum_{i=1}^{r} n_i = 1 + \deg(D).$$

□

Example 2.38 Look at the curve of Examples 2.21 and 2.33. We saw that $f = x/(y + z)$ and $g = y/(y + z)$ are regular outside Q and have a pole of order 2 and 3, respectively, in $Q = (0 : 1 : 1)$. So the functions 1, f and g have mutually distinct pole orders and are elements of $\mathcal{L}(3Q)$. Hence the dimension of $\mathcal{L}(3Q)$ is at least 3. We will see in Example 2.57 that it is exactly 3.

2.5 Differentials on a curve

Let $\mathcal{X}$ be an irreducible smooth curve with function field $\mathbb{F}(\mathcal{X})$.

Definition 2.39 Let $\mathcal{V}$ be a vector space over $\mathbb{F}(\mathcal{X})$. An $\mathbb{F}$-linear map $D : \mathbb{F}(\mathcal{X}) \to \mathcal{V}$ is called a *derivation* if it satifies the *product rule*

$$D(fg) = fD(g) + gD(f).$$

Example 2.40 Let $\mathcal{X}$ be the projective line with funtion field $\mathbb{F}(X)$. Define $D(F) = \sum ia_i X^{i-1}$ for a polynomial $F = \sum a_i X^i \in \mathbb{F}[X]$ and extend this definition to quotients by

$$D\left(\frac{F}{G}\right) = \frac{GD(F) - FD(G)}{G^2}.$$

Then $D : \mathbb{F}(X) \to \mathbb{F}(X)$ is a derivation.

Definition 2.41 The set of all derivations $D : \mathbb{F}(\mathcal{X}) \to \mathcal{V}$ will be denoted by $Der(\mathcal{X}, \mathcal{V})$. We denote $Der(\mathcal{X}, \mathcal{V})$ by $Der(\mathcal{X})$ if $\mathcal{V} = \mathbb{F}(\mathcal{X})$.

The sum of two derivations $D_1, D_2 \in Der(\mathcal{X}, \mathcal{V})$ is defined by $(D_1 + D_2)(f) = D_1(f) + D_2(f)$. The product of $D \in Der(\mathcal{X}, \mathcal{V})$ with $f \in \mathbb{F}(\mathcal{X})$ is defined by $(fD)(g) = fD(g)$. In this way $Der(\mathcal{X}, \mathcal{V})$ becomes a vector space over $\mathbb{F}(\mathcal{X})$.

Theorem 2.42 *Let t be a local parameter at a point P. Then there exists a unique derivation $D_t : \mathbb{F}(\mathcal{X}) \to \mathbb{F}(\mathcal{X})$ such that $D_t(t) = 1$. Furthermore $Der(\mathcal{X})$ is one dimensional over $\mathbb{F}(\mathcal{X})$ and D_t is a basis element for every local parameter t.*

Definition 2.43 A *rational differential form* or *differential* on $\mathcal{X}$ is an $\mathbb{F}(\mathcal{X})$-linear map from $Der(\mathcal{X})$ to $\mathbb{F}(\mathcal{X})$. The set of all rational differential forms on $\mathcal{X}$ is denoted by $\Omega(\mathcal{X})$.

Again $\Omega(\mathcal{X})$ becomes a vector space over $\mathbb{F}(\mathcal{X})$ in the obvious way. Consider the map

$$d : \mathbb{F}(\mathcal{X}) \longrightarrow \Omega(\mathcal{X}),$$

where for $f \in \mathbb{F}(\mathcal{X})$ the differential $df : Der(\mathcal{X}) \to \mathbb{F}(\mathcal{X})$ is defined by $df(D) = D(f)$ for all $D \in Der(\mathcal{X})$. Then d is a derivation.

Theorem 2.44 *The space $\Omega(\mathcal{X})$ has dimension 1 over $\mathbb{F}(\mathcal{X})$ and dt is a basis for every point P with local parameter t.*

So for every point P and local parameter t_P, a differential ω can be represented in a unique way as $\omega = f_P \, dt_P$, where f_P is a rational function. The obvious definition for "the value " of ω in P by $\omega(P) = f_P(P)$ has no meaning, since it depends on the choice of t_P. Despite of this negative result it is possible to say whether ω has a pole or a zero at P of a certain order.

Definition 2.45 Let ω be a differential on $\mathcal{X}$. The *order* or *valuation* of ω in P is defined by $\mathrm{ord}_P(\omega) = v_P(\omega) = v_P(f_P)$. The differential form ω is called *regular* if it has no poles. The regular differentials on $\mathcal{X}$ form an $\mathbb{F}[\mathcal{X}]$-module, which we denote by $\Omega[\mathcal{X}]$.

This definition does not depend on the choices made.

If $\mathcal{X}$ is an affine plane curve defined by the equation $F = 0$ with $F \in \mathbb{F}[X, Y]$, then $\Omega[\mathcal{X}]$ is generated by dx and dy as an $\mathbb{F}[\mathcal{X}]$-module with the relation $f_x dx + f_y dy = 0$.

Example 2.46 We again look at the curve $\mathcal{X}$ in $\mathbb{P}^2$ given by $X^3+Y^3+Z^3 = 0$ in characteristic unequal to three. We define the sets U_x by $U_x = \{(x : y : z) \in \mathcal{X} \mid y \neq 0, z \neq 0\}$ and similarly U_y and U_z. Then U_x, U_y, and U_z cover $\mathcal{X}$ since there is no point on $\mathcal{X}$ where two coordinates are zero. It is easy to check that the three representations

$$\omega = \left(\frac{y}{z}\right)^2 d\left(\frac{x}{y}\right) \text{ on } U_x, \quad \eta = \left(\frac{z}{x}\right)^2 d\left(\frac{y}{z}\right) \text{ on } U_y, \quad \zeta = \left(\frac{x}{y}\right)^2 d\left(\frac{z}{x}\right) \text{ on } U_z$$

define *one* differential on $\mathcal{X}$. For instance, to show that η and ζ agree on $U_y \cap U_z$ one takes the equation $(x/z)^3 + (y/z)^3 + 1 = 0$, differentiates, and applies the formula $d(f^{-1}) = -f^{-2} \, df$ to $f = z/x$.

The only regular functions on $\mathcal{X}$ are constants, so one cannot represent this differential as $g \, df$ with f and g regular functions on $\mathcal{X}$.

Now the divisor of a differential is defined as for functions.

Definition 2.47 The divisor (ω) of the differential ω is defined by

$$(\omega) = \sum_{P \in \mathcal{X}} v_P(\omega) P.$$

Of course, one must show that only finitely many coefficients in (ω) are not 0.

Let ω be a differential and $W = (\omega)$. Then W is called a *canonical divisor*. If ω' is another nonzero differential, then $\omega' = f\omega$ for some rational function f. So $(\omega') = W' \equiv W$ and therefore the canonical divisors form one equivalence class. This class is also denoted by W. Now consider the space $\mathcal{L}(W)$. This space of rational functions can be mapped onto an isomorphic space of differential forms by $f \mapsto f\omega$. By the definition of $\mathcal{L}(W)$, the image of f under the mapping is a regular differential form, that is to say, $\mathcal{L}(W)$ is isomorphic to $\Omega[X]$.

Definition 2.48 Let $\mathcal{X}$ be a smooth projective curve over $\mathbb{F}$. We define the *genus* g of $\mathcal{X}$ by $g = l(W)$.

Example 2.49 Consider the differential dx on the projective line. Then dx is regular at all points $P_a = (a : 1)$, since $x - a$ is a local parameter in P_a and $dx = d(x - a)$. Let $Q = (1 : 0)$ be the point at infinity. Then $t = 1/x$ is a local parameter in Q and $dx = -t^{-2}dt$. So $v_Q(dx) = -2$. Hence $(dx) = -2Q$ and $l(-2Q) = 0$. Therefore the projective line has genus zero.

The genus of a curve will play an important role in the following sections. For methods with which one can determine the genus of a curve, we must refer to textbooks on algebraic geometry. We mention one formula without proof, the so-called *Plücker formula*.

Theorem 2.50 *If $\mathcal{X}$ is a nonsingular projective curve of degree m in $\mathbb{P}^2$, then*

$$g = \frac{1}{2}(m-1)(m-2).$$

Example 2.51 The genus of a line and a nonsingular conic are zero by Theorem 2.50. In fact a curve of genus zero is isomorphic to the projective line. For example the curve $\mathcal{X}$ with equation $XZ - Y^2 = 0$ of Example 2.9 is isomorphic to $\mathbb{P}^1$ where the isomorphism is given by $(x : y : z) \mapsto (x : y) = (y : z)$ for $(x : y : z) \in \mathcal{X}$. The inverse map is given by $(u : v) \mapsto (u^2 : uv : v^2)$.

Example 2.52 So the curve of Examples 2.21, 2.33 and 2.46 has genus 1 and by the definition of genus, $\mathcal{L}(W) = \mathbb{F}$, so regular differentials on $\mathcal{X}$ are scalar multiples of the differential ω of Example 2.46.

For the construction of codes over algebraic curves that generalize Goppa codes, we shall need the concept of residue of a differential at a point P. This is defined in accordance with our treatment of local behavior of a differential ω.

Definition 2.53 Let P be a point on $\mathcal{X}$, t a local parameter at P and $\omega = f\,dt$ the representation of ω. The function f can be written as $\sum_i a_i t^i$. We define the *residue* $\mathrm{Res}_P(\omega)$ of ω in the point P to be a_{-1}.

One can show that this algebraic definition of the residue does not depend on the choice of the local parameter t.

One of the basic results in the theory of algebraic curves is known as the *residue theorem*. We only state the theorem.

Theorem 2.54 *If ω is a differential on a smooth projective curve $\mathcal{X}$, then*

$$\sum_{P \in X} Res_P(\omega) = 0.$$

2.6 The Riemann-Roch theorem

The following theorem, known as the *Riemann-Roch theorem* is not only a central result in algebraic geometry with applications in other areas, but it is also the key to the new results in coding theory.

Theorem 2.55 *Let D be a divisor on a smooth projective curve of genus g. Then, for any canonical divisor W*

$$l(D) - l(W - D) = deg(D) - g + 1.$$

We do not give the proof. The theorem allows us to determine the degree of canonical divisors.

Corollary 2.56 *For a canonical divisor W, we have $deg(W) = 2g - 2$.*

Proof. Everywhere regular functions on a projective curve are constant, that is to say, $\mathcal{L}(0) = \mathbb{F}$, so $l(0) = 1$. Substitute $D = W$ in Theorem 2.55 and the result follows from Definition 2.48. □

Example 2.57 It is now clear why in Example 2.38 the space $\mathcal{L}(3Q)$ has dimension 3. By Example 2.52 the curve $\mathcal{X}$ has genus 1, the degree of $W-3Q$ is negative, so $l(W-3Q) = 0$. By Theorem 2.55 we have $l(3Q) = 3$.

At first, Theorem 2.55 does not look too useful. However, Corollary 2.56 provides us with a means to use it successfully.

Corollary 2.58 *Let D be a divisor on a smooth projective curve of genus g and let $deg(D) > 2g-2$. Then*

$$l(D) = deg(D) - g + 1.$$

Proof. By Corollary 2.56, $\deg(W - D) < 0$, so by Theorem 2.37(i), $l(W - D) = 0$. □

Example 2.59 Consider the code of Theorem 2.27. We embed the affine plane in a projective plane and consider the rational functions on the curve defined by G. By Bézout's theorem, this curve intersects the line at infinity, that is to say, the line defined by $Z = 0$, in m points. These are the possible poles of our rational functions, each with order at most l. So, in the terminology of Definition 2.36, we have a space of rational functions, defined by a divisor D of degree lm. Then Corollary 2.58 and Theorem 2.27 imply that the curve defined by G has genus at most equal to $\binom{m-1}{2}$. This is exactly what we find from the Plücker formula 2.50.

Let m be a nonnegative integer. Then $l(mP) \leq l((m-1)P) + 1$, by the argument as in the proof of Theorem 2.37.

Definition 2.60 If $l(mP) = l((m-1)P)$, then m is called a *(Weierstrass) gap* of P. A nonnegative integer that is not a gap is called a *nongap* of P.

The number of gaps of P is equal to the genus g of the curve, since $l(iP) = i + 1 - g$ if $i > 2g - 2$, by Corollary 2.58 and

$$1 = l(0) \leq l(P) \leq \cdots \leq l((2g-1)P) = g.$$

If $m \in \mathbb{N}_0$, then m is a nongap of P if and only if there exists a rational function which has a pole of order m in P and no other poles. Hence, if m_1 and m_2 are nongaps of P, then $m_1 + m_2$ is also a nongap of P. The nongaps form the *Weierstrass semigroup* in $\mathbb{N}_0$. Let $(\rho_i | i \in \mathbb{N})$ be an enumeration of all the nongaps of P in increasing order, so $\rho_1 = 0$. Let $f_i \in L(\rho_i P)$ be such that $v_P(f_i) = -\rho_i$ for $i \in \mathbb{N}$. Then $f_1, \ldots, f_i$ provide a basis for the space $\mathcal{L}(\rho_i P)$. This will be the approach of Sections 3-7.

The term $l(W - D)$ in Theorem 2.55 can be interpreted in terms of differentials. We introduce a generalization of Definition 2.36 for differentials.

Definition 2.61 Let D be a divisor on a curve $\mathcal{X}$. We define

$$\Omega(D) = \{\omega \in \Omega(\mathcal{X}) \,|\, (\omega) - D \succcurlyeq 0\}$$

and we denote the dimension of $\Omega(D)$ over $\mathbb{F}$ by $\delta(D)$, called the *index of speciality* of D.

The connection with functions is established by the following theorem.

Theorem 2.62 *$\delta(D) = l(W - D)$.*

Proof. If $W = (\omega)$, we define a linear map $\phi : \mathcal{L}(W - D) \to \Omega(D)$ by $\phi(f) = f\omega$. This is clearly an isomorphism. □

Example 2.63 If we take $D = 0$, then by Definition 2.48 there are exactly g linearly independent regular differentials on a curve $\mathcal{X}$. So the differential of Example 2.46 is the only regular differential on $\mathcal{X}$ (up to a constant factor) as was already observed after Theorem 2.50.

2.7 Codes from algebraic curves

We now come to the applications to coding theory. Our alphabet will be $\mathbb{F}_q$. Let $\mathbb{F}$ be the algebraic closure of $\mathbb{F}_q$. We shall apply the theorems of the previous sections. A few adaptations are necessary, since for example, we consider for functions in the coordinate ring only those that have coefficients in $\mathbb{F}_q$. If the affine curve $\mathcal{X}$ over $\mathbb{F}_q$ is defined by a prime ideal I in $\mathbb{F}_q[X_1, \ldots, X_n]$, then its coordinate ring $\mathbb{F}_q[\mathcal{X}]$ is by definition equal to $\mathbb{F}_q[X_1, \ldots, X_n]/I$ and its function field $\mathbb{F}_q(\mathcal{X})$ is the quotient field of $\mathbb{F}_q[\mathcal{X}]$.

It is always assumed that the curve is *absolutely irreducible*, this means that the the defining ideal is also prime in $\mathbb{F}[X_1, \ldots, X_n]$. Similar adaptions are made for projective curves. Notice that $F(x_1, \ldots, x_n)^q = F(x_1^q, \ldots, x_n^q)$ for all $F \in \mathbb{F}_q[X_1, \ldots, X_n]$. So if $(x_1, \ldots, x_n)$ is a zero of F and F is defined over $\mathbb{F}_q$, then $(x_1^q, \ldots, x_n^q)$ is also a zero of F. Let $Fr : \mathbb{F} \to \mathbb{F}$ be the *Frobenius map* defined by $Fr(x) = x^q$. We can extend this map coordinatewise to points in affine and projective space. If $\mathcal{X}$ is a curve defined over $\mathbb{F}_q$ and P is a point of $\mathcal{X}$, then $Fr(P)$ is also a point of $\mathcal{X}$, by the above remark. A divisor D on $\mathcal{X}$ is called rational if the coefficients of P and $Fr(P)$ in D are the same for any point P of $\mathcal{X}$. The space $\mathcal{L}(D)$ will only be considered for rational divisors and is defined as before but with the restriction of the rational functions to $\mathbb{F}_q(\mathcal{X})$. With these changes the stated theorems remain true over $\mathbb{F}_q$ in particular the theorem of Riemann-Roch 2.55.

Let $\mathcal{X}$ be an absolutely irreducible nonsingular projective curve over $\mathbb{F}_q$. We shall define two kinds of algebraic geometry codes from $\mathcal{X}$. The first kind generalizes Reed-Solomon codes, the second kind generalizes Goppa codes. In the following, $P_1, P_2, \ldots, P_n$ are rational points on $\mathcal{X}$ and D is the divisor $P_1 + P_2 + \cdots + P_n$. Furthermore G is some other divisor that has support disjoint from D. Although it is not necessary to do so, we shall make more restrictions on G, namely

$$2g - 2 < \deg(G) < n.$$

Definition 2.64 The linear code $C(D, G)$ of length n over $\mathbb{F}_q$ is the image of the linear map $\alpha : \mathcal{L}(G) \to \mathbb{F}_q^n$ defined by $\alpha(f) = (f(P_1), f(P_2), \ldots, f(P_n))$. Codes of this kind are called *geometric Reed Solomon codes.*

Theorem 2.65 *The code $C(D, G)$ has dimension $k = deg(G) - g + 1$ and minimum distance $d \geq n - deg(G)$.*

Proof. (i) If f belongs to the kernel of α, then $f \in \mathcal{L}(G - D)$ and by Theorem 2.37(i), this implies $f = 0$. The result follows from the assumption $2g - 2 < \deg(G) < n$ and Corollary 2.58.

(ii) If $\alpha(f)$ has weight d, then there are $n-d$ points P_i, say $P_{i_1}, P_{i_2}, \ldots, P_{i_{n-d}}$, for which $f(P_i) = 0$. Therefore $f \in \mathcal{L}(G - E)$, where $E = P_{i_1} + \cdots + P_{i_{n-d}}$. Hence $\deg(G) - n - d \geq 0$. □

Note the analogy with the proof of Theorem 2.27.

Example 2.66 Let $\mathcal{X}$ be the projective line over $\mathbb{F}_{q^m}$. Let $n = q^m - 1$. We define $P_0 = (0 : 1)$, $P_\infty = (1 : 0)$ and we define the divisor D as $\sum_{j=1}^{n} P_j$, where $P_j = (\beta^j : 1)$, $(1 \leq j \leq n)$. We define $G = aP_0 + bP_\infty$, $a \geq 0$, $b \geq 0$. (Here β is a primitive nth root of unity.) By Theorem 2.55, $\mathcal{L}(G)$ has dimension $a + b + 1$ and one immediately sees that the functions $(x/y)^i$, $-a \leq i \leq b$, form a basis of $\mathcal{L}(G)$. Consider the code $C(D, G)$. A generator matrix for this code has as rows $(\beta^i, \beta^{2i}, \ldots, \beta^{ni})$ with $-a \leq i \leq b$. One easily checks that $(c_1, c_2, \ldots, c_n)$ is a codeword in $C(D, G)$ if and only if $\sum_{j=1}^{n} c_j(\beta^l)^j = 0$ for all l with $a < l < n - b$. It follows that $C(D, G)$ is a Reed-Solomon code. The subfield subcode with coordinates in $\mathbb{F}_q$ is a BCH code.

Example 2.67 Let $\mathcal{X}$ be the curve of Examples 2.21, 2.33, 2.38 and 2.57. Let $G = 3Q$, where $Q = (0 : 1 : 1)$. We take $n = 8$, so D is the sum of the remaining rational points. The coordinates are given by

	Q	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
x	0	0	0	1	α	$\overline{\alpha}$	1	α	$\overline{\alpha}$
y	1	α	$\overline{\alpha}$	0	0	0	1	1	1
z	1	1	1	1	1	1	0	0	0

where $\overline{\alpha} = \alpha^2 = 1 + \alpha$. We saw in Examples 2.38 and 2.57 that 1, $x/(y + z)$ and $y/(y + z)$ are a basis of $\mathcal{L}(3Q)$ over $\mathbb{F}$ and hence also over $\mathbb{F}_4$. This leads to the following generator matrix for $C(D, G)$:

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

By Theorem 2.65, the minimum distance is at least 5 and of course, one immediately sees from the generator matrix that $d = 5$.

We now come to the second class of algebraic geometry codes. We shall call these codes *geometric Goppa codes.*

Definition 2.68 The linear code $C^*(D, G)$ of length n over $\mathbb{F}_q$ is the image of the linear map $\alpha^* : \Omega(G - D) \rightarrow \mathbb{F}_q^n$ defined by

$$\alpha^*(\eta) = (\mathrm{Res}_{P_1}(\eta), \mathrm{Res}_{P_2}(\eta), \ldots, \mathrm{Res}_{P_n}(\eta)).$$

The parameters are given by the following theorem.

Theorem 2.69 *The code $C^*(D,G)$ has dimension $k^* = n - deg(G) + g - 1$ and minimum distance $d^* \geq deg(G) - 2g + 2$.*

Proof. Just as in Theorem 2.65, these assertions are direct consequences of Theorem 2.55 (Riemann-Roch), using Theorem 2.62 (making the connection between the dimension of $\Omega(G)$ and $l(W - G)$) and Corollary 2.56 (stating that the degree of a canonical divisor is $2g - 2$). □

Example 2.70 Let $L = \{\alpha_1, \ldots, \alpha_n\}$ be a set of n distinct elements of $\mathbb{F}_{q^m}$. Let g be a polynomial in $\mathbb{F}_{q^m}[X]$ which is not zero at α_i for all i. The *(classical) Goppa code* $\Gamma(L,g)$ is defined by

$$\Gamma(L,g) = \{\mathbf{c} \in \mathbb{F}_q^n \mid \sum \frac{c_i}{X - \alpha_i} \equiv 0 \ (mod\ g\ )\}.$$

Let $P_i = (\alpha_i : 1)$, $Q = (1 : 0)$ and $D = P_1 + \cdots + P_n$. If we take for E the divisor of zeros of g on the projective line, then $\Gamma(L,g) = C^*(D, E - Q)$ and

$$\mathbf{c} \in \Gamma(L,g) \text{ if and only if } \sum \frac{c_i}{X - \alpha_i} dX \in \Omega(E - Q - D).$$

This is the reason that some authors extend the definiton of geometric Goppa codes to subfield subcodes of codes of the form $C^*(D,G)$.

It is a well-known fact that the parity check matrix of the Goppa code $\Gamma(L,g)$ is equal to the following generator matrix of a generalized RS code

$$\begin{pmatrix} g(\alpha_1)^{-1} & \ldots & g(\alpha_n)^{-1} \\ \alpha_1 g(\alpha_1)^{-1} & \ldots & \alpha_n g(\alpha_n)^{-1} \\ \vdots & \ldots & \vdots \\ \alpha_1^{r-1} g(\alpha_1)^{-1} & \ldots & \alpha_n^{r-1} g(\alpha_n)^{-1} \end{pmatrix},$$

where r is the degree of the Goppa polynomial g. So $\Gamma(L,g)$ is the subfield subcode of the dual of a generalized RS code. This is a special case of the following theorem.

Theorem 2.71 *The codes $C(D,G)$ and $C^*(D,G)$ are dual codes.*

Proof. From Theorem 2.65 and Theorem 2.69 we know that $k + k^* = n$. So it suffices to take a word from each code and show that the inner product of the two words is 0. Let $f \in \mathcal{L}(G)$, $\eta \in \Omega(G - D)$. By Definitions 2.64 and

2.68, the differential $f\eta$ has no poles except possibly poles of order 1 in the points $P_1, P_2, \dots, P_n$. The residue of $f\eta$ in P_i is equal to $f(P_i)\mathrm{Res}_{P_i}(\eta)$. By Theorem 2.54, the sum of the residues of $f\eta$ over all the poles, that is to say, over the points P_i, is equal to zero. Hence we have

$$0 = \sum_{i=1}^{n} f(P_i)\mathrm{Res}_{P_i}(\eta) = \langle \alpha(f), \alpha^*(\eta) \rangle.$$

□

Several authors prefer the codes $C^*(D, G)$ over geometric RS codes but the nonexperts in algebraic geometry probably feel more at home with polynomials than with differentials. That this is possible without loss of generality is stated in the following theorem.

Theorem 2.72 *Let $\mathcal{X}$ be a curve defined over $\mathbb{F}_q$. Let $P_1, \dots, P_n$ be n rational points on $\mathcal{X}$. Let $D = P_1 + \cdots + P_n$. Then there exists a differential form ω with simple poles at the P_i such that $Res_{P_i}(\omega) = 1$ for all i. Furthermore*

$$C^*(D, G) = C(D, W + D - G)$$

for all divisors G that have a support disjoint from the support of D, where W is the divisor of ω.

So one can do without differentials and the codes $C^*(D, G)$. However, it is useful to have both classes when treating decoding methods. These use parity checks, so one needs a generator matrix for the dual code.

In the next paragraph we treat several examples of algebraic geometry codes. It is already clear that we find some *good* codes. For example from Theorem 2.65 we see that such codes over a curve of genus 0 (the projective line) are MDS codes. In fact, Theorem 2.65 says that $d \geq n - k + 1 - g$, so if g is small, we are close to the Singleton bound.

2.8 Some algebraic geometry codes

We know that to find good codes, we must find long codes. To use the methods from algebraic geometry, it is necessary to find rational points on a given curve. The number of these is a bound on the length of the code. A central problem in algebraic geometry is finding bounds for the number of rational points on a variety. In order to appreciate some of the examples in this paragraph, we mention without proof the improvement by *Serre* of the *Hasse-Weil bound.*

Theorem 2.73 *Let $\mathcal{X}$ be a curve of genus g over $\mathbb{F}_q$. If $N_q(\mathcal{X})$ denotes the number of rational points on $\mathcal{X}$, then*

$$|N_q(\mathcal{X}) - (q+1)| \leq g\lfloor 2\sqrt{q} \rfloor.$$

Example 2.74 In this example we consider codes from *Hermitian* curves of Example 2.13. Let $q = r^2$. Consider the second affine equation $X^{r+1} = Y^r + Y$ of the Hermitian curve $\mathcal{X}$ in $\mathbb{A}^2$ over $\mathbb{F}_q$. By Theorem 2.50, the genus g of $\mathcal{X}$ equals $\frac{1}{2}r(r-1) = \frac{1}{2}(q - \sqrt{q})$. We shall first show that $\mathcal{X}$ has the maximal number of rational points, that is to say, by Theorem 2.73 exactly $1 + q\sqrt{q}$. The last equation has $(0:1:0)$ as the only point at infinity. To see that the number of affine $\mathbb{F}_q$-rational points is $r + (r+1)(r^2 - r) = r^3$ one argues as follows. The right side of the equation $X^{r+1} = Y^r + Y$ is the trace from $\mathbb{F}_q$ to $\mathbb{F}_r$. The first r in the formula on the number of points corresponds to the elements of $\mathbb{F}_r$. These are exactly the elements of $\mathbb{F}_q$ with zero trace. The remaining term corresponds to the elements in $\mathbb{F}_q$ with a nonzero trace, since the equation $X^{r+1} = \beta$, $\beta \in \mathbb{F}_r^*$, has exactly $r+1$ solutions in $\mathbb{F}_q$.

We take $G = mQ$, where $Q = (0:1:0)$ and $q - \sqrt{q} < m < q\sqrt{q}$. The code $C(D,G)$ over $\mathbb{F}_q$ has length $n = q\sqrt{q}$, dimension $k = m - g + 1$, and distance $d \geq n - m$. We will deal with the true minimum distance in Section 5.3. To see how good these codes are, we take as example $q = 16$. A basis for $\mathcal{L}(G)$ is easily found. The functions $f_{i,j} = x^i y^j / z^{i+j}$, $0 \leq i \leq 4$, $4i + 5j \leq m$ will do the job. First, observe that there are $m - 5 = m - g + 1$ pairs (i,j) satisfying these conditions. The functions x/z and y/z can be treated in exactly the same way as in Examples 2.33 and 2.34, showing that $f_{i,j}$ has a pole of order $4i + 5j$ in Q. Hence, these functions are independent. Therefore, the code is easily constructed. Decoding will be treated in Sections 6 and 7. Let us try to get some idea of the quality of this code. Suppose that we intend to send a long message (say 10^9 bits) over a channel with an error probability $p_e = 0.01$ (quite a bad channel). We compare coding using a rate $\frac{1}{2}$ Reed-Solomon code over $\mathbb{F}_{16}$ with using $C(D,G)$, where we take $m = 37$ to also have rate $\frac{1}{2}$. In this case, $C(D,G)$ has distance 27. The RS code has word length 16 (so 64 bits) and distance 9. If a word is received incorrectly, we assume that all the bits are wrong when we count the number of errors. For the RS code, the error probability after decoding is roughly $3 \cdot 10^{-4}$; however, for the code $C(D,G)$, the error probability after decoding is less than $2 \cdot 10^{-7}$. In this example, it is important to keep in mind that we are fixing the alphabet (in this case $\mathbb{F}_{16}$). If we compare the code $C(D,G)$, for which the words are strings of 256 bits, with a rate $\frac{1}{2}$ RS code over $\mathbb{F}_{2^5}$ (words are 160 bits long), the latter will come close in performance (error probability $2 \cdot 10^{-6}$) and a rate $\frac{1}{2}$ RS code over $\mathbb{F}_{2^6}$ (words are 384 bits long) performs better (roughly 10^{-7}).

One could also compare our code with a binary BCH code of length 255 and rate about $\frac{1}{2}$. The BCH code wins when we are concerned with random errors. If we are using a bursty channel, then the code $C(D,G)$ can handle bursts of length up to 46 bits (which influence at most 13 letters of a codeword) while the BCH code would fail completely.

Example 2.75 Let $\mathcal{X}$ be the Klein quartic over $\mathbb{F}_8$ of Example 2.20. By Theorem 2.50, the genus is 3. By Theorem 2.73, $\mathcal{X}$ can have at most 24 rational points and as we saw in Example 2.20 it has 24 rational points. Let $Q = (0:0:1)$ and let D be the sum of the other 23 rational points, $G = 10Q$. From Theorem 2.65, we find that $C(D,G)$ has dimension $10 - g + 1 = 8$ and minimum distance $d \geq 23 - 10 = 13$. We now concatenate this code with the [4,3,2] single parity check code as follows. The symbols in codewords of $C(D,G)$ are elements of $\mathbb{F}_8$ which we interpret as column vectors of length 3 over $\mathbb{F}_2$ and then we adjoin the parity check. The resulting code C is a binary $[92, 24, 26]$ code. The punctured code, a $[91, 24, 25]$ code set a new world record for codes with $n = 91$, $d = 25$.

Example 2.76 We show how to construct a generator matrix for the code of the previous example. We consider the functions x/z and y/z. The divisors $(x/z) = 3P_1 - P_2 - 2Q$ and $(y/z) = P_1 + 2P_2 - 3Q$ were computed in Example 2.34. From these divisors, we can deduce that the functions $(x/z)^i(y/z)^j$ with $0 \leq 2i + 3j \leq 10$, $0 \leq i \leq 2j$ are in $\mathcal{L}(10Q)$. We thus have eight functions in $\mathcal{L}(10Q)$ with poles in Q of order 0,3,5,6,7,8,9, and 10, respectively. Hence they are independent and since $l(10Q) = 8$, they are a basis of $\mathcal{L}(10Q)$. By substituting the coordinates of the rational points of $\mathcal{X}$ in these functions, we find the 8 by 23 generator matrix of the code.

Example 2.77 Let $\mathbb{F}_4 = \{0, 1, \alpha, \overline{\alpha}\}$, where $\alpha^2 = \alpha + 1 = \overline{\alpha}$. Consider the curve $\mathcal{X}$ over $\mathbb{F}_4$ given by the equation $x^2y + \alpha y^2 z + \overline{\alpha} z^2 x = 0$. This is a nonsingular curve with genus 1. Its nine rational points are given by

	P_1	P_2	P_3	P_4	P_5	P_6	Q_1	Q_2	Q_3
x	1	0	0	1	1	1	α	1	1
y	0	1	0	α	$\overline{\alpha}$	1	1	α	1
z	0	0	1	$\overline{\alpha}$	α	1	1	1	α

Let $D = P_1 + P_2 + \cdots + P_6$, $G = 2Q_1 + Q_2$. We claim that the functions $x/(x + y + \overline{\alpha}z)$, $y/(x + y + \overline{\alpha}z)$, $\overline{\alpha}z/(x + y + \overline{\alpha}z)$ are a basis of $\mathcal{L}(G)$. To see this, note that the numerators in these fractions are not 0 in Q_1 and Q_2 and that the line with equation $x + y + \overline{\alpha}z = 0$ meets $\mathcal{X}$ in Q_2 and is tangent to $\mathcal{X}$ in Q_1. By Theorem 2.65, the code $C(D, G)$ of length 6 has minimum distance at least 3. However, the code is in fact an MDS code, namely the *hexacode.*

2.9 Asymptotically good sequences of codes and curves

The parameters of a linear block code over the finite field $\mathbb{F}_q$ of *length* n, *dimension* k and *minimum distance* d will be denoted by $[n, k, d]_q$ or $[n, k, d]$. The quotient k/n is called the *information rate* and denoted by $R = k/n$ and the *relative minimum distance* d/n is denoted by δ.

The dimension k and the minimum distance d of an algebraic geometry code on a curve of genus g with n points that are defined over $\mathbb{F}_q$ satisfy

$$k + d \geq n + 1 - g,$$

by Theorem 2.65. Hence

$$R + \delta \geq 1 - \frac{g - 1}{n}.$$

Definition 2.78 A sequence of codes $(C_m | m \in \mathbb{N})$ with parameters $[n_m, k_m, d_m]$ over a fixed finite field $\mathbb{F}_q$ is called *asymptotically good* if n_m tends to infinity, and d_m/n_m tends to a nonzero constant δ, and k_m/n_m tends to a nonzero constant R for $m \to \infty$.

Let $H_q(0) = 0$ and $H_q(x) = x\log_q(q-1) - x\log_q x - (1-x)\log_q(1-x)$ for $0 < x \leq (q-1)/q$ be the entropy function. Then there exist asymptotically good sequences of codes attaining the the *Gilbert-Varshamov* bound

$$R \geq 1 - H_q(\delta).$$

In order to construct asymptotically good codes we therefore need curves with low genus and many $\mathbb{F}_q$-rational points.

Definition 2.79 Let $N_q(g)$ be the maximal number of $\mathbb{F}_q$-rational points on an absolutely irreducible nonsingular projective curve over $\mathbb{F}_q$ of genus g. Let

$$A(q) = \limsup_{g\to\infty} \frac{N_q(g)}{g}.$$

The Hasse-Weil bound 2.73 implies

$$A(q) \leq 2\sqrt{q}.$$

This has been improved to the *Drinfeld-Vlăduţ* bound.

Theorem 2.80

$$A(q) \leq \sqrt{q} - 1.$$

Furthermore equality holds if q is a square.

The equality is proved by studying the number of rational points of *modular curves* over finite fields. The theory of modular curves is a central and very important part of mathematics, but it is very involved and deep, much more so than the theory concerning the Riemann-Roch theorem and we will not touch it.

Applying this to algebraic geometry codes one derives the following *Tsfasman-Vlăduţ-Zink* (TVZ) bound.

Theorem 2.81 *Let q be a square. Then for every R there exists an asymptotically good sequence of codes such that the limit value of the information rate is R and the relative minimum distance is δ and*

$$R + \delta \geq 1 - \frac{1}{\sqrt{q} - 1}.$$

This in turn means that the TVZ bound is better than the GV bound when q is a square and $q \geq 49$ in a certain range of δ. This fact was the starting point of the current interest in algebraic geometry codes.

In the following we discuss an alternative method to derive these results. Let F be a polynomial in the variables X and Y with coefficients in $\mathbb{F}_q$. Let $a = \deg_Y(F)$. Assume that there exists a subset S of $\mathbb{F}_q$ such that for any given $x \in S$ there exist exactly a distinct $y_1, \ldots, y_a \in S$ such that $F(x, y_i) = 0$ for all $i = 1, \ldots, a$. Consider the algebraic set $\mathcal{X}_m$ in $\mathbb{A}^m$ defined by the equations

$$F(X_i, X_{i+1}) = 0 \text{ for } i = 1, \ldots, m-1.$$

A lower bound on the number of rational points of $\mathcal{X}_m$ is easily seen to be $\#S \cdot a^{m-1}$ by induction. If $\mathcal{X}_m$ is absolutely irreducible, then it is a curve.

Example 2.82 Let $F = (X^q - X) - (Y^q - Y)$. Then F is an example with $a = q$ and $S = \mathbb{F}_q$. Then $\mathcal{X}_m$ has q^m rational points. This is the maximal possible number of rational points for an algebraic set in $\mathbb{A}^m$, but $\mathcal{X}_m$ is reducible, since F is divisible by $X - Y$.

Example 2.83 Let $F = X(X^q - X) - (Y^q - Y)$. Then F is an example with $a = q$ and $S = \mathbb{F}_q$. One can show that $\mathcal{X}_m$ is a curve. The number of rational points of $\mathcal{X}_m$ is again q^m, but the genus of these curves grows faster than the number of rational points.

Definition 2.84 A sequence of curves $(\mathcal{X}_m | m \in \mathbb{N})$ is called *asymptotically good* if $g(\mathcal{X}_m)$ tends to infinity and the following limit exists and

$$\lim_{m \to \infty} \frac{N_q(\mathcal{X}_m)}{g(\mathcal{X}_m)} > 0,$$

where $g(\mathcal{X})$ is the genus of $\mathcal{X}$ and $N_q(\mathcal{X})$ is the number of $\mathbb{F}_q$-rational points of $\mathcal{X}$.

Example 2.85 Let $q = 8$. Let $F = XY^3 + Y + X^3$. Then F is an example with $a = 3$ and $S = \mathbb{F}_8^*$. Therefore this gives a curve with $7 \cdot 3^{m-1}$ points with nonzero coordinates in $\mathbb{F}_8$, but this sequence of curves is not asymptotically good.

Example 2.86 Let $q = 4$. Let $F = XY^2 + Y + X^2$. Then F is an example with $a = 2$ and $S = \mathbb{F}_4^*$. Therefore this gives a curve with $3 \cdot 2^{m-1}$ points with nonzero coordinates in $\mathbb{F}_4$, and in fact it gives a sequence of curves that is asymptotically good.

More generally, let $q = r^2$ and consider $F = X^{r-1}Y^r + Y - X^r$. Then we get an example with $a = r$ and $S = \mathbb{F}_q^*$, that is to say, the equation $F = 0$ has the property that for every given nonzero element $x \in \mathbb{F}_q$ there are exactly r nonzero solutions in $\mathbb{F}_q$ of the equation $F(x, Y) = 0$ in Y. This is seen by multiplying the equation by X and replacing XY by Z. Then the equation $Z^r + Z = X^{r+1}$ is obtained, which defines the Hermitian curve over $\mathbb{F}_q$, which we have considered before in Example 2.74. Therefore the corresponding sequence of curves $\mathcal{X}_m$ satifies

$$N_q(\mathcal{X}_m) \geq (q-1)r^{m-1}.$$

The genus of the curve $\mathcal{X}_m$ is computed by induction by applying the formula of *Hurwitz-Zeuthen* to the covering $\pi_m : \mathcal{X}_m \to \mathcal{X}_{m-1}$, where π_m is defined as $\pi_m(x_1, \ldots, x_m) = (x_1, \ldots, x_{m-1})$. In this case it turns out to be an *Artin-Schreier covering.* It is easier to view this in terms of function fields. Let $\mathcal{F}_m$ be the function field of $\mathcal{X}_m$. Then $\mathcal{F}_1 = \mathbb{F}_q(z_1)$ and $\mathcal{F}_m$ is obtained from $\mathcal{F}_{m-1}$ by adjoining a new element z_m that satisfies the equation

$$z_m^r + z_m = x_{m-1}^{r+1},$$

where $x_{m-1} = z_{m-1}/x_{m-2} \in \mathcal{F}_{m-1}$ for $m \geq 2$, and $x_1 = z_1$, $x_0 = 1$.

Theorem 2.87 *The genus g_m of the curve $\mathcal{X}_m$, or equivalently of the function field $\mathcal{F}_m$ is equal to*

$$g_m = \begin{cases} r^m + r^{m-1} - r^{\frac{m+1}{2}} - 2r^{\frac{m-1}{2}} + 1 & \textit{if } \ m \textit{ is odd }, \\ r^m + r^{m-1} - \frac{1}{2}r^{\frac{m+2}{2}} - \frac{3}{2}r^{\frac{m}{2}} - r^{\frac{m-2}{2}} + 1 & \textit{if } \ m \textit{ is even }. \end{cases}$$

Thus the Drinfeld-Vlăduţ bound is attained.

It turns out that finding bases for the vector spaces involved in the construction of AG codes is difficult. This last part remains to be done in order to make the codes really constructive.

A new sequence of curves $\mathcal{Y}_m$ with function field $\mathcal{T}_m$ over $\mathbb{F}_q$ with $q = r^2$ is given as follows. Let $\mathcal{T}_1 = \mathbb{F}_q(X_1)$. Let $\mathcal{T}_m$ be obtained from $\mathcal{T}_{m-1}$ by adjoining a new element x_m that satisfies the equation:

$$x_m^r + x_m = \frac{x_{m-1}^r}{x_{m-1}^{r-1} + 1}.$$

By induction it is shown that

$$N_q(\mathcal{Y}_m) \geq (r^2 - r)r^{m-1}.$$

The same method applies to derive the following theorem.

Theorem 2.88 *The genus g_m of the curve $\mathcal{Y}_m$ is equal to*

$$g_m = \begin{cases} (r^{\frac{m+1}{2}} - 1)(r^{\frac{m-1}{2}} - 1) & \textit{if } \ m \textit{ is odd },\\ (r^{\frac{m}{2}} - 1)^2 & \textit{if } \ m \textit{ is even }. \end{cases}$$

Hence this sequence of function fields attains the Drinfeld-Vlăduţ bound too.

Let Q_m be the rational point on the curve $\mathcal{Y}_m$ that is the unique pole of x_1.

Theorem 2.89 *Let Λ_m be the Weierstrass semigroup of Q_m. Then $\Lambda_1 = \mathbb{N}_0$ and*

$$\Lambda_{m+1} = r \cdot \Lambda_m \cup \{n \in \mathbb{N}_0 \mid n \geq c_m\},$$

where

$$c_m = \begin{cases} r^m - r^{\frac{m+1}{2}} & \textit{if } \ m \textit{ is odd },\\ r^m - r^{\frac{m}{2}} & \textit{if } \ m \textit{ is even }. \end{cases}$$

This means that the sequence of nongaps of Q_m is known, but an explicit description of a basis for the spaces $\mathcal{L}(iQ_m)$ is not known in general.

2.10 Notes

Goppa submitted his seminal paper [37] in June 1975 and it was published in 1977. Goppa also published three more papers in the eighties [38, 39, 40] and a book [41] in 1991.

The material treated in Sections 2.1-2.8 can be found in the textbooks [62, 70, 97, 100] and the survey [61]. For books on algebraic geometry we refer to [2, 13, 14, 31, 44, 90, 104, 105] to mention a few.

The codes on plane curves in Theorem 2.27 using Bézout's theorem are a special cases of Goppa's construction and come from [50]. The Hermitian curves in Example 2.13 and their codes have been studied by many authors. See [91, 95, 96, 99, 106]. The Klein curve goes back to F. Klein [54] and has been studied thoroughly, also over finite fields in connection with codes. See [17, 19, 25, 43, 46, 62].

The world record mentioned in Example 2.75 is taken from [8]. More results in this direction are mentioned in [11].

A survey on bounds on the number of rational points on curves and its relation with coding theory one can find in [11, 34]. The upper bound on $A(q)$ in Theorem 2.80 was shown in [16]. The equality was proved in [48, 101], see also [58, 100].

The construction of the modular curves and the corresponding codes can be done with polynomial complexity, of degree 20 for classical modular curves and degree 30 for Drinfeld modular curves, see [66, 100]. The degree for the latter has been reduced to 17 in [64].

The first negative result on asymptotically good sequences of curves is in [30]. The computation of the genus of the curves in Example 2.83 is from [77]. Example 2.85 and the idea to construct asymptotically good codes in this way is from [27] and in [36] it is shown that this sequence of curves is not asymptotically good. The two sequences of asymtotically good curves as presented in Theorems 2.87 and 2.88 are from [35, 36].

The computation of the Weierstrass semigroups of Theorem 2.89 can be found in [78].

A first step in the direction of comuting the spaces $\mathcal{L}(iQ)$ is made in [103] for the curve $\mathcal{X}_3$ and in [42] for $\mathcal{X}_4$ and $q = 16$.

Conference proceedings concerning AG codes are [76, 98] and a special issue on this topic appeared in [59].

3 Order functions

The construction of codes in Section 2.3 can be generalized to the so-called evaluation codes as will be done in Section 4. To this end we introduce the notions of order, degree and weight functions and treat a method to obtain such functions.

3.1 Order, degree, and weight functions

Recall that an $\mathbb{F}$-*algebra* is a commutative ring with a unit that contains $\mathbb{F}$ as a unitary subring. $\mathbb{N}$ denotes the positive integers and $\mathbb{N}_0$ the nonnegative integers.

The standard example of an $\mathbb{F}$-algebra is $R = \mathbb{F}[X_1, \ldots, X_m]$. To present the codes we need an order of a special kind on the polynomials in R which we define as follows.

Definition 3.1 Let $R = \mathbb{F}[X_1, \ldots, X_m]$. Suppose that $\prec$ is a total order on the set of monomials in the variables $X_1, \ldots, X_m$ such that for all monomials M_1, M_2, and M, the following hold

$$\begin{array}{ll} (R.1) & \text{If } M \neq 1 \text{, then } 1 \prec M, \\ (R.2) & \text{If } M_1 \prec M_2 \text{, then } MM_1 \prec MM_2. \end{array}$$

Then $\prec$ is called a *reduction* order, *term* order or *admissible* order on the monomials.

The multi-index notation will be used for monomials. That means $X^\alpha = \prod_{i=1}^m X_i^{\alpha_i}$ if $\alpha = (\alpha_1, \ldots, \alpha_m)$. The degree of a monomial is defined by

$$\deg(X^\alpha) = \deg(\alpha) = \sum_{i=1}^m \alpha_i.$$

Giving a reduction order on monomials in m variables is the same as giving a total order on $\mathbb{N}_0^m$ such that, for all α_1, α_2, and α in $\mathbb{N}_0^m$, the following hold

$$\begin{array}{ll} (E.1) & \text{If } \alpha \neq 0 \text{, then } 0 \prec \alpha, \\ (E.2) & \text{If } \alpha_1 \prec \alpha_2 \text{, then } \alpha + \alpha_1 \prec \alpha + \alpha_2. \end{array}$$

We use $\prec$ both for monomials and exponents.

Example 3.2 The *lexicographic order* $\prec_L$ is defined by

$$X^\alpha \prec_L X^\beta \text{ if and only if}$$
$$\alpha_1 = \beta_1, \ldots, \alpha_{l-1} = \beta_{l-1} \text{ and } \alpha_l < \beta_l \text{ for some } l,\ 1 \leq l \leq m.$$

The lexicographic order is a reduction order. For $m = 2$, with $X = X_1$, $Y = X_2$ and $\prec = \prec_L$, the lexicographic order looks like

$$\begin{array}{lllllllllll}
1 & \prec & Y & \prec & Y^2 & \prec & \cdots & \prec & Y^j & \prec & Y^{j+1} & \prec & \cdots \\
X & \prec & XY & \prec & XY^2 & \prec & \cdots & \prec & XY^j & \prec & XY^{j+1} & \prec & \cdots \\
X^2 & \prec & \cdots
\end{array}$$

So X^{i+1} is the supremum of the set $\{ X^iY^j \mid j \in \mathbb{N}_0 \}$. If $m \geq 2$, then the lexicographic order is not isomorphic with the positive integers with the usual order.

Example 3.3 The *graded lexicographic order* $\prec_D$ is defined by

$$X^\alpha \prec_D X^\beta \text{ if and only if}$$
$$\text{either } \deg(X^\alpha) < \deg(X^\beta) \text{ or } \deg(X^\alpha) = \deg(X^\beta) \text{ and } X^\alpha \prec_L X^\beta.$$

The graded lexicographic order is a reduction order which is isomorphic with the positive integers with the usual order.

An order is extended to a function on all polynomials in the following way. Let $\prec$ be a reduction order which is isomorphic with the positive integers with the usual order. Let $f_1, f_2, \ldots$ be the enumeration of the set of monomials such that $f_i \prec f_{i+1}$ for all i. The monomials constitutes a basis of $\mathbb{F}[X_1, \ldots, X_m]$ over $\mathbb{F}$. So every nonzero polynomial f can be written in a unique way as

$$f = \sum_{i=1}^{j} \lambda_i f_i,$$

where $\lambda_i \in \mathbb{F}$ for all i, and $\lambda_j \neq 0$. Define a function

$$\rho : \mathbb{F}[X_1, \ldots, X_m] \longrightarrow \mathbb{N}_0 \cup \{-\infty\},$$

by $\rho(0) = -\infty$ and $\rho(f) = j - 1$ where j is the smallest positive integer such that f can be written as a linear combination of the first j monomials. It is not difficult to show that ρ satisfies the following conditions

$(O.0)$ $\rho(f) = -\infty$ if and only if $f = 0$

$(O.1)$ $\rho(\lambda f) = \rho(f)$ for all nonzero $\lambda \in \mathbb{F}$

$(O.2)$ $\rho(f + g) \leq \max\{\rho(f), \rho(g)\}$
and equality holds when $\rho(f) < \rho(g)$.

$(O.3)$ If $\rho(f) < \rho(g)$ and $h \neq 0$, then $\rho(fh) < \rho(gh)$

$(O.4)$ If $\rho(f) = \rho(g)$, then there exists a nonzero $\lambda \in \mathbb{F}$ such that
$\rho(f - \lambda g) < \rho(g)$.

for all $f, g, h \in R$. Here $-\infty < n$ for all $n \in \mathbb{N}_0$. The properties of the function ρ are captured in the following definition.

Definition 3.4 Let R be an $\mathbb{F}$-algebra. An *order function* on R is a map

$$\rho : R \longrightarrow \mathbb{N}_0 \cup \{-\infty\},$$

that satisfies the conditions $(O.0), \ldots, (O.4)$.

Definition 3.5 Let R be an $\mathbb{F}$-algebra. A *weight function* on R is an order function on R that satisfies furthermore

$$(O.5) \quad \rho(fg) = \rho(f) + \rho(g)$$

for all $f, g \in R$. Here $-\infty + n = -\infty$ for all $n \in \mathbb{N}_0$.

If ρ is a weight function and $\rho(f)$ is divisible by an integer $d > 1$ for all $f \in R$, then $\rho(f)/d$ is again a weight function. So we may assume that the greatest common divisor of the integers $\rho(f)$ with $0 \neq f \in R$ is 1.

Definition 3.6 A *degree function* on R is a map that satisfies conditions $(O.0)$, $(O.1)$, $(O.2)$ and $(O.5)$.

It is clear that condition $(O.3)$ is a consequence of $(O.5)$.

Example 3.7 The standard example of an $\mathbb{F}$-algebra R with a degree function ρ is obtained by taking $R = \mathbb{F}[X_1, \ldots, X_m]$ and $\rho(f) = \deg(f)$, the degree of $f \in R$. It is an order function if and only if $m = 1$, and here it is also a weight function.

Example 3.8 Let $\mathcal{X}$ be an absolutely irreducible nonsingular projective curve over field $\mathbb{F}$. Let P be an $\mathbb{F}$-rational point. R be the ring of rational functions on $\mathcal{X}$ that have no poles outside P. So, if $f \in R$ and $v_P(f) \geq 0$, then f is regular on $\mathcal{X}$ and therefore constant. Hence $v_P(f) \leq 0$ for all nonzero $f \in R$. Define $\rho(f) = -v_P(f)$ for $f \in R$. Then ρ is a weight function by Theorem 2.16. The purpose of this and the following sections is to show that it is possible to develop the theory of AG codes to a certain extent without the theory of algebraic curves.

Lemma 3.9 *Let ρ be an order function on R. Then we have:*
(1) *If $\rho(f) = \rho(g)$, then $\rho(fh) = \rho(gh)$ for all $h \in R$.*
(2) *If $f \in R$ and $f \neq 0$, then $\rho(1) \leq \rho(f)$.*
(3) *$\mathbb{F} = \{ f \in R \mid \rho(f) \leq \rho(1) \}$.*
(4) *If $\rho(f) = \rho(g)$, then there exists a unique nonzero $\lambda \in \mathbb{F}$ such that $\rho(f - \lambda g) < \rho(g)$.*

Proof.

(1) Let $\rho(f) = \rho(g)$. Then $(O.4)$ says that there exists a nonzero $\lambda \in \mathbb{F}$ such that $\rho(f - \lambda g) < \rho(g)$. So $\rho(fh - \lambda gh) < \rho(gh)$, by $(O.3)$. Now $fh = (fh - \lambda gh) + \lambda gh$. So $\rho(fh) = \rho(\lambda gh) = \rho(gh)$, by $(O.2)$ and $(O.1)$, respectively.

(2) Suppose that f is a nonzero element of R such that $\rho(f) < \rho(1)$. Then $\rho(1) > \rho(f) > \rho(f^2) > \cdots$ is a strictly decreasing sequence, by condition $(O.3)$, but this contradicts the fact that $\mathbb{N}_0 \cup \{-\infty\}$ is a well-order. Hence $\rho(1) \leq \rho(f)$ for all nonzero elements f in R.

(3) It is clear that $\mathbb{F}$ is a subset of $\{ f \in R \mid \rho(f) \leq \rho(1) \}$, by conditions $(O.0)$ and $(O.1)$. If f is nonzero and $\rho(f) \leq \rho(1)$, then $\rho(f) = \rho(1)$, by (2). Hence there exists a nonzero $\lambda \in \mathbb{F}$ such that $\rho(f - \lambda 1) < \rho(1)$, by $(O.4)$. So $f - \lambda = 0$ and $f \in \mathbb{F}$.

(4) The existence is guaranteed by $(O.4)$. For the uniqueness we argue as follows. Suppose that there exist nonzero $\lambda, \mu \in \mathbb{F}$ such that $\rho(f - \lambda g) < \rho(g)$ and $\rho(f - \mu g) < \rho(g)$. We get by $(O.1)$ and $(O.2)$ that $\rho(f - \lambda g - (f - \mu g)) < \rho(g)$. Therefore $\rho((\mu - \lambda)g) < \rho(g)$. Condition $(O.1)$ gives $\mu - \lambda = 0$. □

Proposition 3.10 *If there exists an order function on R, then R is an integral domain.*

Proof. Suppose that $fg = 0$ for some nonzero $f, g \in R$. We may assume that $\rho(f) \leq \rho(g)$. So $\rho(f^2) \leq \rho(fg) = \rho(0) = -\infty$. So $\rho(f^2) = -\infty$, and $f^2 = 0$. Now $f \neq 0$, hence $\rho(1) \leq \rho(f)$, by Lemma 3.9. So $\rho(f) \leq \rho(f^2) = \rho(0) = -\infty$. Hence $f = 0$, which is a contradiction. Therefore R has no zero divisors. □

Example 3.11 The $\mathbb{F}$-algebra $R = \mathbb{F}[X_1, X_2]/(X_1X_2 - 1)$ is an integral domain. We will show that it does not have an order function. Denote the coset of X_i modulo the ideal $(X_1X_2 - 1)$ by x_i. If ρ is an order function on R, then $\rho(1) \leq \rho(x_1)$, so $\rho(x_2) \leq \rho(x_1x_2) = \rho(1)$. Hence $\rho(x_2) = \rho(1)$ and in the same way we get $\rho(x_1) = \rho(1)$. Therefore $\rho(f) \leq \rho(1)$ for all $f \in R$. So $\mathbb{F} = R$ by Lemma 3.9, which is a contradiction since $x_1 \notin \mathbb{F}$.

The following proposition and theorem show that if there exists an order function, then there exists a basis with certain properties; and conversely if such a basis exists, then one can define an order function. Although the formulation is technical, it is easy to apply. This will be shown in some examples.

Proposition 3.12 *Let R be an $\mathbb{F}$-algebra with order function ρ. Assume that $R \neq \mathbb{F}$. Then there exists a basis $\{ \ f_i \mid i \in \mathbb{N} \ \}$ of R over $\mathbb{F}$ such that $\rho(f_i) < \rho(f_{i+1})$ for all i. Every such basis has the property that if i is the smallest positive integer such that f can be written as a linear combination of the first i elements of that basis, then $\rho(f) = \rho(f_i)$.Let $l(i,j)$ be the integer l such that $\rho(f_i f_j) = \rho(f_l)$, then $l(i,j) < l(i+1,j)$ for all i and j. Let $\rho_i = \rho(f_i)$. If ρ is a weight function, then $\rho_{l(i,j)} = \rho_i + \rho_j$.*

Proof. There exists an $f \in R$ such that $f \notin \mathbb{F}$, since $R \neq \mathbb{F}$. So $\rho(1) < \rho(f)$ by Lemma 3.9. Hence $\rho(f^n) < \rho(f^{n+1})$ for all $n \in \mathbb{N}_0$. Therefore the set of values of ρ is infinite. Let $(\rho_i \mid i \in \mathbb{N})$ be the increasing sequence of all nonnegative integers that appear as the order $\rho(f)$ of a nonzero element $f \in R$. By definition for all $i \in \mathbb{N}$ there exists an $f_i \in R$ such that $\rho(f_i) = \rho_i$. So $\rho(f_i) < \rho(f_{i+1})$ for all i, and for all nonzero $f \in R$ there exists an i with $\rho(f) = \rho(f_i)$, by definition. The fact that $\{ \ f_i \mid i \in \mathbb{N} \ \}$ is a basis is proved by induction and Lemma 3.9 (4), and it has the required property by $(O.2)$. That the function $l(i,j)$ is strictly increasing in its first argument is a consequence of condition $(O.3)$. If ρ is a weight function, then $\rho_{l(i,j)} = \rho_i + \rho_j$ by condition $(O.5)$. □

Example 3.13 Consider the graded lexicographic order as in Example 3.3 for $m = 2$ with $X = X_1$ and $Y = X_2$. Let R be the $\mathbb{F}$-algebra $\mathbb{F}[X,Y]$. It has

$$\{ \ X^\alpha Y^\beta \mid \alpha, \beta \in \mathbb{N}_0 \ \}$$

as basis. Consider the basis of monomials and their corresponding indexing in the following two dimensional arrays:

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
Y^6	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdots$
Y^5	XY^5	$\cdot$	$\cdot$	$\cdot$	$\cdots$
Y^4	XY^4	X^2Y^4	$\cdot$	$\cdot$	$\cdots$
Y^3	XY^3	X^2Y^3	X^3Y^3	$\cdot$	$\cdots$
Y^2	XY^2	X^2Y^2	X^3Y^2	$\cdot$	$\cdots$
Y	XY	X^2Y	X^3Y	X^4Y	$\cdots$
1	X	X^2	X^3	X^4	X^5

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
22	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdots$
16	23	$\cdot$	$\cdot$	$\cdot$	$\cdots$
11	17	24	$\cdot$	$\cdot$	$\cdots$
7	12	18	25	$\cdot$	$\cdots$
4	8	13	19	$\cdot$	$\cdots$
2	5	9	14	20	$\cdots$
1	3	6	10	15	21

So the enumeration is along the diagonals from left above to right below. Consider the elements $f_8 = XY^2$ and $f_9 = X^2Y$. Then

$$f_8 f_9 = XY^2 \cdot X^2Y = X^3Y^3 = f_{25}.$$

Hence $l(8,9) = 25$.

Theorem 3.14 *Let R be an $\mathbb{F}$-algebra. Let $\{ f_i \mid i \in \mathbb{N} \}$ be a basis of R as a vector space over $\mathbb{F}$ with $f_1 = 1$. Let L_i be the vector space generated by $f_1, \ldots, f_i$. Let $l(i,j)$ be the smallest positive integer l such that $f_i f_j \in L_l$. Suppose $l(i,j) < l(i+1,j)$ for all $i,j \in \mathbb{N}$. Let $(\rho_i \mid i \in \mathbb{N})$ be a strictly increasing sequence of nonnegative integers. Define $\rho(0) = -\infty$, and $\rho(f) = \rho_i$ if i is the smallest positive integer such that $f \in L_i$. Then ρ is an order function on R. If moreover $\rho_{l(i,j)} = \rho_i + \rho_j$, then ρ is a weight function.*

Proof. The conditions $(O.0)$, $(O.1)$, $(O.2)$ and $(O.4)$ are a direct consequence of the definitions.

With every nonzero element $f \in R$ we associate $\iota(f)$, the smallest positive integer such that $f \in L_{\iota(f)}$. Let f and g be nonzero elements of R. Then

$$f = \sum_{i \leq \iota(f)} \lambda_i f_i, \ g = \sum_{j \leq \iota(g)} \nu_j f_j \ \text{ and } \ fg = \sum_{l \leq \iota(fg)} \mu_l f_l,$$

with $\lambda_{\iota(f)} \neq 0$, $\nu_{\iota(g)} \neq 0$ and $\mu_{\iota(fg)} \neq 0$. There exist $\mu_{ijl} \in \mathbb{F}$ such that

$$f_i f_j = \sum_{l \leq l(i,j)} \mu_{ijl} f_l$$

and $\mu_{ijl(i,j)} \neq 0$. So

$$\mu_l = \sum_{l(i,j)=l} \lambda_i \nu_j \mu_{ijl}.$$

The function $l(i,j)$ is strictly increasing in both arguments, by assumption and symmetry. So $l(i,j) < l(\iota(f), \iota(g))$ if $i < \iota(f)$ or $j < \iota(g)$. Furthermore, if $i = \iota(f)$ and $j = \iota(g)$, then

$$\lambda_i \nu_j \mu_{ijl(i,j)} \neq 0,$$

This element is therefore equal to $\mu_{\iota(fg)}$, and we have proved that $\iota(fg) = l(\iota(f), \iota(g))$.

If moreover $\rho_{l(i,j)} = \rho_i + \rho_j$, then

$$\rho(fg) = \rho_{\iota(fg)} = \rho_{l(\iota(f),\iota(g))} = \rho_{\iota(f)} + \rho_{\iota(g)} = \rho(f) + \rho(g).$$

□

Example 3.15 Let $\mathbf{w} = (w_1, \ldots, w_m)$ be an m-tuple of positive integers called *weights.* The *weighted degree* of $\alpha \in \mathbb{N}_0^m$ and the corresponding monomial X^α is defined by

$$\text{wdeg}(X^\alpha) = \text{wdeg}(\alpha) = \sum \alpha_l w_l,$$

and of a nonzero polynomial $F = \sum \lambda_\alpha X^\alpha$ by

$$\text{wdeg}(F) = \max\{ \text{ wdeg}(X^\alpha) \mid \lambda_\alpha \neq 0 \ \}.$$

This gives a degree function wdeg on the ring $\mathbb{F}[X_1, \ldots, X_m]$. The *weighted graded lexicographic order* $\prec_{\mathbf{w}}$ on $\mathbb{N}_0^m$ is defined by

$$\begin{gathered} \alpha \prec_{\mathbf{w}} \beta \ \text{ if and only if} \\ \text{either wdeg}(\alpha) < \text{wdeg}(\beta) \text{ or wdeg}(\alpha) = \text{wdeg}(\beta) \text{ and } \alpha \prec_L \beta \ , \end{gathered}$$

and similarly for the monomials. This is indeed a reduction order that is isomorphic to $\mathbb{N}$.

Consider the weighted graded lexicographic order for $m = 2$ with $X = X_1$, $Y = X_2$, $\text{wdeg}(X) = 4$ and $\text{wdeg}(Y) = 5$. Let R be the $\mathbb{F}$-algebra $\mathbb{F}[X, Y]$. It has $\{ \ X^\alpha Y^\beta \mid \alpha, \beta \in \mathbb{N}_0 \ \}$ as basis. Consider the weighted degrees $4\alpha + 5\beta$ of this basis and their corresponding indexing in the following two dimensional arrays:

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
25	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdots$
20	24	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdots$
15	19	23	$\cdot$	$\cdot$	$\cdot$	$\cdots$
10	14	18	22	$\cdot$	$\cdot$	$\cdots$
5	9	13	17	21	25	$\cdots$
0	4	8	12	16	20	24

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
22	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdots$
15	20	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdots$
10	14	19	$\cdot$	$\cdot$	$\cdot$	$\cdots$
6	9	13	18	$\cdot$	$\cdot$	$\cdots$
3	5	8	12	17	23	$\cdots$
1	2	4	7	11	16	21

The basis elements X^6 and XY^4 have both 24 as weighted degree. But XY^4 is smaller than X^6 in the lexicographic order. Hence $f_{20} = XY^4$ and $f_{21} = X^6$.

3.2 Existence of weight functions

Example 3.16 Let I be the ideal in $\mathbb{F}[X,Y]$ generated by a polynomial of the form

$$X^aY^c + Y^{b+c} + G$$

with $G \in \mathbb{F}[X,Y]$, $\deg_X(G) = d < a$, $\deg(G) < b+c$ and $\gcd(a,b) = 1$, where the degree of $G \in \mathbb{F}[X,Y]$ as a polynomial in X is denoted by $\deg_X(G)$. Let $S = \mathbb{F}[X,Y]/I$. Denote the cosets of X, Y and G modulo I by x, y and g, respectively. Then $x^ay^c = -y^{b+c} - g$ and therefore x^ay^c is a linear combination of elements of the form $x^\alpha y^\beta$ with $\alpha < a$, since $\deg_X(G) < a$. One shows by induction that the set

$$\{\ x^\alpha y^\beta \mid \alpha, \beta \in \mathbb{N}_0, \alpha < a \text{ or } \beta < c\ \}$$

is a basis for S. Suppose that there exists a weight function ρ on S such that $\gcd(\rho(x), \rho(y)) = 1$. We will show that $\rho(x) = b$ and $\rho(y) = a$. Let $X^\alpha Y^\beta$ be the monomial in G with the largest weight. Then $\alpha + \beta < b + c$. So $\rho(g) \leq \alpha\rho(x) + \beta\rho(y)$ by $(O.2)$ and $(O.5)$.

(1) If $\rho(y) \leq \rho(x)$, then

$$\alpha\rho(x) + \beta\rho(y) = \alpha\rho(x) + (\beta - c)\rho(y) + c\rho(y) \leq$$

$$(\alpha + \beta - c)\rho(x) + c\rho(y) < a\rho(x) + c\rho(y).$$

So $\rho(g) < \rho(x^a y^c)$. Hence $\rho(x^a y^c) = \rho(x^a y^c + g)$. But $\rho(y^{b+c}) = \rho(x^a y^c + g)$. Therefore $\rho(y^{b+c}) = \rho(x^a y^c)$. So $a\rho(x) = b\rho(y)$.

(2) If $\rho(x) \leq \rho(y)$, then by a similar argument we conclude that $a\rho(x) = b\rho(y)$.

Hence in both cases $a\rho(x) = b\rho(y)$. But $\gcd(\rho(x), \rho(y)) = 1$. So $\rho(x) = b$ and $\rho(y) = a$.

We will see in the following proposition that the $\mathbb{F}$-algebra S has such a weight function if $c = 0$. But if $c > 0$, then x^a and y^b are two elements that have the same weight ab and are independent modulo elements of weight strictly smaller than ab. This contradicts condition $(O.4)$. Therefore there is no weight function if $c > 0$.

The polynomial $X^3Y + Y^3 + Y$ is reducible and is of the above form with $a = 3$, $b = 2$, $c = 1$, $d = 0$ and $G = Y$. So by Proposition 3.10 an order function does not exist.

Consider the subspace R of S that is generated by

$$\{\ x^\alpha y^\beta \mid \alpha, \beta \in \mathbb{N}_0, \alpha < a \text{ and } c\alpha \leq (a-d)\beta\ \}.$$

In the following it is shown that R is an $\mathbb{F}$-algebra and that indeed a weight function exists on R such that $\rho(x) = b$ and $\rho(y) = a$. The choice may seem ad hoc, but the affine curve with equation $X^aY^c + Y^{b+c} + G = 0$ has the points $P = (1:0:0)$ and $Q = (0:1:0)$ at infinity if $c > 0$. By computing the divisors of the monomials $x^i y^j$ as explained in Section 2 one shows that the ring R consists of all functions in S that are also regular in P, so that have possibly a pole in Q and nowhere else.

Proposition 3.17 *Let I be the ideal in $\mathbb{F}[X,Y]$ generated by a polynomial of the form $X^aY^c + uY^{b+c} + G$ with $u \in \mathbb{F}^*$, $G \in \mathbb{F}[X,Y]$, $deg_X(G) = d < a$, $deg(G) < b + c$ and $\gcd(a,b) = 1$. Let $S = \mathbb{F}[X,Y]/I$. Let R be the vector space generated by $\{\ x^\alpha y^\beta \mid \alpha, \beta \in \mathbb{N}_0, \alpha < a$ and $c\alpha \leq (a-d)\beta\ \}$. Then R is an $\mathbb{F}$-algebra with a weight function ρ such that $\rho(x) = b$ and $\rho(y) = a$.*

Proof. The set $\{\ x^\alpha y^\beta \mid \alpha < a \text{ or } \beta < c\ \}$ is a basis for S over $\mathbb{F}$. So $\{\ x^\alpha y^\beta \mid \alpha < a \text{ and } c\alpha \leq (a-d)\beta\ \}$ is a basis for R. Let $f_1, f_2, \ldots$ be an enumeration of this basis of R. If $f_i = x^\alpha y^\beta$, $\alpha < a$ and $c\alpha \leq (a-d)\beta$, then define $\rho_i = \alpha b + \beta a$. The map $(\alpha, \beta) \mapsto \alpha b + \beta a$ is injective on the domain $\{\ (\alpha, \beta) \in \mathbb{N}_0^2 \mid \alpha < a\ \}$, since $\gcd(a,b) = 1$. Therefore if $i \neq j$, then $\rho_i \neq \rho_j$.

So we may assume that the enumeration is such that $(\rho_i | i \in \mathbb{N}_0)$ is a strictly increasing sequence.

Let $L_l = \langle f_1, \ldots, f_l \rangle$. We will prove that for all i, j there exists a nonnegative integer l such that $f_i f_j \in L_l$. So R is an $\mathbb{F}$-algebra. Furthermore we will show that if $l(i,j)$ is the smallest nonnegative integer l such that $f_i f_j \in L_l$, then $\rho_{l(i,j)} = \rho_i + \rho_j$. Hence there exists a weight function ρ on R such that $\rho(x^\alpha y^\beta) = \alpha b + \beta a$, by Theorem 3.14.

Let $f_i = x^\alpha y^\beta$, $\rho_i = \alpha b + \beta a$ with $\alpha < a$ and $c\alpha \leq (a-d)\beta$. Let $f_j = x^\gamma y^\delta$, $\rho_j = \gamma b + \delta a$ with $\gamma < a$ and $c\gamma \leq (a-d)\delta$. Then $f_i f_j = x^{\alpha+\gamma} y^{\beta+\delta}$, $\rho_i + \rho_j = (\alpha+\gamma)b + (\beta+\delta)a$ and $c(\alpha+\gamma) \leq (a-d)(\beta+\delta)$.

(1) If $\alpha + \gamma < a$, then $f_i f_j$ is a basis element of R. So $f_{l(i,j)} = f_i f_j$ and $\rho_{l(i,j)} = \rho_i + \rho_j$.

(2) If $\alpha + \gamma \geq a$, then $\alpha + \gamma = a + \epsilon$ with $\epsilon < a$. Now

$$ca \leq c(\alpha+\gamma) \leq (a-d)(\beta+\delta).$$

So $\beta + \delta = c + \eta$ for some nonnegative integer η; and $c(a+\epsilon) \leq (a-d)(c+\eta)$. Hence $c(d+\epsilon) \leq (a-d)\eta$. So $\epsilon < a$ and $c\epsilon \leq (a-d)(b+c+\eta)$. Furthermore

$$f_i f_j = x^a y^c x^\epsilon y^\eta = -u x^\epsilon y^{b+c+\eta} - x^\epsilon y^\eta g.$$

So the term $x^\epsilon y^{b+c+\eta}$ is a basis element f_l of R, and

$$\rho_i + \rho_j = (\alpha+\gamma)b + (\beta+\delta)a = \epsilon b + (b+c+\eta)a = \rho_l.$$

We will show that $x^\epsilon y^\eta g \in L_{l-1}$. This implies that $f_i f_j \in L_l$ and $f_i f_j \notin L_{l-1}$. So $l(i,j) = l$. A monomial of G with a nonzero coefficient, is of the form $X^\kappa Y^\lambda$ with $\kappa \leq d$ and $\kappa + \lambda < b + c$, since $\deg_X(G) = d$ and $\deg(G) < b + c$. We will prove by induction on ϵ that:

$$\text{If } (\epsilon, \eta),\ (\kappa, \lambda) \in \mathbb{N}_0^2,\ \epsilon < a,\ c(\epsilon + d) \leq (a-d)\eta,\ \ \kappa \leq d,$$

$$\kappa + \lambda < b + c \text{ and } \rho_l = \epsilon b + (b+c+\eta)a, \text{ then } x^{\epsilon+\kappa} y^{\eta+\lambda} \in L_{l-1}.$$

As a result we get that $x^\epsilon y^\eta g \in L_{l-1}$.

(2.i) If $\epsilon + \kappa < a$, then $x^{\epsilon+\kappa} y^{\eta+\lambda}$ is a basis element of R, since $c(\epsilon+\kappa) \leq (a-d)(\eta+\lambda)$. Furthermore

$$(\epsilon+\kappa)b + (\eta+\lambda)a = \epsilon b + (\kappa b + \lambda a) + \eta a < \epsilon b + (b+c+\eta)a = \rho_l,$$

since $b < a$ and $\kappa + \lambda < b + c$. So $x^{\epsilon+\kappa} y^{\eta+\lambda} \in L_{l-1}$.

$(2.ii)$ If $\epsilon+\kappa \geq a$, then $\epsilon+\kappa = a+\epsilon'$, for some nonnegative integer ϵ'. Now $\epsilon' < \epsilon$, since $\kappa \leq d < a$. Similarly as before we have that $\eta+\lambda = c+\eta'$ for some nonnegative integer η', and $c\epsilon' \leq (a-d)\eta'$. Furthermore

$$x^{\epsilon+\kappa}y^{\eta+\lambda} = x^a y^c x^{\epsilon'} y^{\eta'} = -ux^{\epsilon'}y^{b+c+\eta'} - x^{\epsilon'}y^{\eta'}g.$$

So the term $x^{\epsilon'}y^{b+c+\eta'}$ is a basis element $f_{l'}$ of R, and

$$\rho_{l'} = \epsilon' b + (b+c+\eta')a = (a+\epsilon')b + (c+\eta')a = (\epsilon+\kappa)b + (\eta+\lambda)a$$

which is strictly smaller than ρ_l, as we have seen in $(2.i)$. So $x^{\epsilon'}y^{\eta'}g \in L_{l'-1}$ by induction, and $l' < l$. Therefore $x^{\epsilon+\kappa}y^{\eta+\lambda} \in L_{l-1}$. □

Corollary 3.18 *Let F be a polynomial of the form $X^aY^b + uY^{b+c} + G$ with $u \in \mathbb{F}^*$, $G \in \mathbb{F}[X,Y]$, $deg_X(G) = d < a$, $deg(G) < b+c$ and $\gcd(a,b) = 1$. If G is not divisible by Y, then F is absolutely irreducible.*

Proof. Suppose that there are polynomials U and V such that $F = UV$. Let u and v be the cosets of U and V, respecitively, in $S = \mathbb{F}[X,Y]/(F)$. Then $uv = 0$. Let R be the subspace of S generated by the elements $x^\alpha y^\beta$ such that $\alpha < a$ and $c\alpha \leq (a-d)\beta$. Then R is an $\mathbb{F}$-algebra with a weight function by Proposition 3.17. Hence R is an integral domain by Proposition 3.10. Consider the two cases:

(i) If $c = 0$, then $R = S$. Hence $u = 0$ or $v = 0$.

(ii) Suppose $c > 0$. By an argument similar to that in the proof of the previous proposition one shows that there are positive integers r and s such that y^ru, $y^sv \in R$. So $y^ru \cdot y^sv = y^{r+s}uv = 0$. Hence $y^ru = 0$ or $y^sv = 0$. If $y^ru = 0$, then $Y^rU \in (F)$. So there exists a polynomial A such that

$$Y^rU = AF = A(X^aY^b + uY^{b+c} + G).$$

F is not divisible by Y, since $c > 0$ and G is not divisible by Y. Therefore A is divisible by Y^r. So $U \in (F)$ and $u = 0$. Similarly $v = 0$ if $y^sv = 0$.

In both cases $u = 0$ or $v = 0$. Hence S is an integral domain, (F) is a prime ideal and F is irreducible. These results still hold after extending the field $\mathbb{F}$ to its algebraic closure. Therefore F is absolutely irreducible. □

Example 3.19 Let $q = r^2$ be an even prime power. Consider the Hermitian curve over $\mathbb{F}_q$ with the affine equation

$$X^{r+1} - Y^r - Y = 0.$$

See Example 2.13. Then it is of the form $X^aY^c + uY^{b+c} + G = 0$ as treated in Proposition 3.17 with $a = r+1$, $b = r$, $c = d = 0$ and $u = -1$, $G = -Y$. We will have a closer look for $r = 4$. Let R be the $\mathbb{F}_{16}$-algebra $\mathbb{F}_{16}[X,Y]/(X^5 - Y^4 - Y)$. It has

$$\{\ x^\alpha y^\beta \mid \alpha < 5\ \}$$

as basis. Then $\rho(x^\alpha y^\beta) = 4\alpha + 5\beta$ gives a weight function on R. Consider the basis of functions and their corresponding weights in the following two dimensional arrays:

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y^4	$\cdot$	$\cdot$	$\cdot$	$\cdot$
y^3	xy^3	$\cdot$	$\cdot$	$\cdot$
y^2	xy^2	x^2y^2	$\cdot$	$\cdot$
y	xy	x^2y	x^3y	x^4y
1	x	x^2	x^3	x^4

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
20	$\cdot$	$\cdot$	$\cdot$	$\cdot$
15	19	$\cdot$	$\cdot$	$\cdot$
10	14	18	$\cdot$	$\cdot$
5	9	13	17	21
0	4	8	12	16

The sequence $(f_l | l \in \mathbb{N})$ is an enumeration of the basis with increasing weight. The first terms are :

$$1,\ x,\ y,\ x^2,\ xy,\ y^2,\ x^3,\ x^2y,\ xy^2,\ y^3,\ x^4,\ x^3y,\ x^2y^2,\ xy^3,\ y^4,\ x^4y,\ \cdots$$

Consider for instance $f_4 = x^2$ and $f_7 = x^3$. Then

$$f_4 f_7 = x^5 = -y^4 - y = -f_{15} - f_3.$$

So $l(4,7) = 15$.

The weights are given by the sequence $(\rho_l | l \in \mathbb{N})$. The weights of the terms above are :

$$0,\ 4,\ 5,\ 8,\ 9,\ 10,\ 12,\ 13,\ 14,\ 15,\ 16,\ 17,\ 18,\ 19,\ 20,\ 21,\ \cdots$$

Hence $\rho_l = l + 5$ for all $l \geq 7$.

Remark 3.20 If $c = 0$, then $R = S$ and $\{\ x^\alpha y^\beta \mid \beta < b\ \}$ is also a basis of the $\mathbb{F}$-algebra, by symmetry.

Example 3.21 The affine equation $X^3Y + Y^3 + X = 0$ over $\mathbb{F}_8$ of the Klein quartic, see Example 2.14, is of the form $X^aY^b + uY^{b+c} + G = 0$ as treated in Proposition 3.17 with $a = 3$, $b = 2$, $c = d = 1$ and $u = 1$, $G = X$. Let R be the $\mathbb{F}_8$-subalgebra of $\mathbb{F}_8[X, Y]/(X^3Y + Y^3 + X)$ generated by the elements $x^\alpha y^\beta$ such that $\alpha < 3$, $\alpha \leq 2\beta$. Then R has

$$\{1\} \cup \{\ x^\alpha y^\beta \mid \alpha \leq 2, 1 \leq \beta\ \}$$

as basis, and $\rho(1) = 0, \rho(x^\alpha y^\beta) = 2\alpha + 3\beta$ gives a weight function on R. Consider the basis of functions and their corresponding weights in the following two dimensional arrays:

$\vdots$	$\vdots$	$\vdots$
y^4	$\cdot$	$\cdot$
y^3	xy^3	x^2y^3
y^2	xy^2	x^2y^2
y	xy	x^2y
1		

$\vdots$	$\vdots$	$\vdots$
12	$\cdot$	$\cdot$
9	11	13
6	8	10
3	5	7
0		

The sequence $(f_l | l \in \mathbb{N})$ is an enumeration of the basis with increasing weight. The first terms are :

$$1,\ y,\ xy,\ y^2,\ x^2y,\ xy^2,\ y^3,\ x^2y^2,\ xy^3,\ y^4,\ x^2y^3,\ \cdots$$

Consider $f_3 = xy$ and $f_5 = x^2y$. Then

$$f_3f_5 = x^3y^2 = -y^4 - xy = -f_{10} - f_3.$$

So $l(3, 5) = 10$.

The weights are given by the sequence $(\rho_l | l \in \mathbb{N})$. The weights of the terms above are :

$$0,\ 3,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 12,\ 13,\ \cdots$$

Hence $\rho_l = l + 2$ for all $l \geq 3$.

Example 3.22 This is partly a specialization to $c = 0$ and partly a generalization to more than two variables of Proposition 3.17. Let wdeg be the weighted degree on $\mathbb{F}[X_1, \ldots, X_m]$, where X_i has weight $a_1 \cdots a_{i-1}b_i \cdots b_{m-1}$. Let I be the ideal in $\mathbb{F}[X_1, \ldots, X_m]$ generated by

$$X_i^{a_i} + X_{i+1}^{b_i} + G_i \text{ for } i = 1, \ldots, m - 1,$$

where $G_i \in \mathbb{F}[X_1, \ldots, X_{i+1}]$, $\mathrm{wdeg}(G_i) < a_1 \cdots a_i b_i \cdots b_{m-1}$ and $\gcd(a_i, b_j) = 1$ for all $i \leq j$. Then the ring $R = \mathbb{F}[X_1, \ldots, X_m]/I$ has

$$\{ \ x^\alpha \mid \alpha \in \mathbb{N}_0^m, \ \alpha_i < a_i \text{ for all } i < m \ \}$$

as a basis. The affine ring R has a weight function ρ such that

$$\rho(x_i) = a_1 \cdots a_{i-1} b_i \cdots b_{m-1}.$$

Therefore the ring R is an integral domain and the ideal I is prime. This can be proved with the theory of Gröbner bases for which we refer to the literature mentioned in the Notes.

3.3 Notes

The approach given in this and the next section is new. One can find the ideas in the work of Feng, Rao, Tzeng and Wei [25, 26, 29], where the notion of an order function is used on the level of codewords instead of functions, with the concept of well-behaving sequences. See also [53, 73, 92].

The notion of a reduction order is standard in the theory of Gröbner bases [12, 14]. In [45, 83] the connection between Gröbner bases and en- and decoding of algebraic geometry codes is treated. The curves of Example 3.16 were considered in [25] and the special case with $c = 0$ in [68, 69].

A similar result to Corollary 3.18 concerning the irreducibility of plane curves is proved by other means in [3].

The generalization of Proposition 3.17 mentioned in Example 3.22 is from [23, 75].

4 Evaluation codes and the dual minimum distance

We define evaluation codes and give bounds on the minimum distance of the dual codes in terms of the order function.

4.1 Evaluation codes and their duals

Let R be an $\mathbb{F}_q$-algebra with an order function ρ. Let $(f_i \mid i \in \mathbb{N})$ be a basis of R over $\mathbb{F}_q$ such that $\rho(f_i) < \rho(f_{i+1})$ for all $i \in \mathbb{N}$, and for all nonzero

$f \in R$ there exists a j with $\rho(f) = \rho(f_j)$. The existence of such a basis is guaranteed by Proposition 3.12. Let L_l be the vector space generated by $f_1, \dots, f_l$. Then for all nonzero $f \in R$ we have that $\rho(f) = \rho(f_l)$ if and only if l is the smallest integer such that $f \in L_l$. Let $l(i,j)$ be the smallest positive integer l such that $f_i f_j \in L_l$. So $l(i,j) < l(i+1,j)$ for all $i, j \in \mathbb{N}$.

The coordinatewise multiplication on $\mathbb{F}_q^n$ is defined by $\mathbf{a}*\mathbf{b} = (a_1b_1, \dots, a_nb_n)$ for $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{b} = (b_1, \dots, b_n)$. The vector space $\mathbb{F}_q^n$ with the multiplication $*$ becomes a commutative ring with the unit $(1, \dots, 1)$. Identify the unitary subring $\{(\lambda, \dots, \lambda) | \lambda \in \mathbb{F}_q\}$ with $\mathbb{F}_q$. In this way $\mathbb{F}_q^n$ is an $\mathbb{F}_q$-algebra .

Definition 4.1 The map

$$\varphi : R \longrightarrow \mathbb{F}_q^n,$$

is called a *morphism* of $\mathbb{F}_q$-algebras if φ is $\mathbb{F}_q$-linear and

$$\varphi(fg) = \varphi(f) * \varphi(g).$$

Let $\mathbf{h}_i = \varphi(f_i)$. Define the *evaluation code* E_l and its dual C_l by

$$E_l = \varphi(L_l) = \langle \mathbf{h}_1, \dots, \mathbf{h}_l \rangle,$$

$$C_l = \{\mathbf{c} \in \mathbb{F}_q^n \mid \mathbf{c} \cdot \mathbf{h}_i = 0 \ \text{ for all } \ i \leq l\}.$$

The sequence of codes $(E_l | l \in \mathbb{N})$ is increasing with respect to the inclusion. They are all subspaces of $\mathbb{F}_q^n$. So there exists an N such that $E_l = E_N$ for all $l \geq N$. The code E_N is the image of R under φ. We will consider only those algebra morphisms φ that are *surjective*. So $E_l = \mathbb{F}_q^n$ and $C_l = 0$ for all $l \geq N$.

Example 4.2 Let the set $\mathcal{P}$ consist of n distinct points $P_1, \dots, P_n$ in $\mathbb{F}_q^m$. Let $R = \mathbb{F}[X_1, \dots, X_m]$. Consider the evaluation map

$$ev_{\mathcal{P}} : R \longrightarrow \mathbb{F}^n,$$

defined by $ev_{\mathcal{P}}(f) = (f(P_1), \dots, f(P_n))$. This is a morphism of $\mathbb{F}_q$-algebras from R to $\mathbb{F}_q^n$, since $FG(P) = F(P)G(P)$ for all polynomials F and G, and all points P.

Lemma 4.3 *The map $ev_{\mathcal{P}}$ is surjective.*

Proof. Let $P_j = (x_{j1}, \ldots, x_{jm})$. Let $A_{il} = \{x_{jl} \mid j = 1, \ldots, n\} \setminus \{x_{il}\}$. Define the polynomial G_i by

$$G_i = \prod_{l=1}^{m} \prod_{x \in A_{il}} (X_l - x).$$

Then $G_i(P_j) = 0$ for all $i \neq j$. Furthermore $G_i(P_i) \neq 0$, since the points $P_1, \ldots, P_n$ are distinct. The polynomial $G_i/G_i(P_i)$ maps under $ev_{\mathcal{P}}$ to the i-th standard basis element of $\mathbb{F}_q^n$. Hence $ev_{\mathcal{P}}$ is surjective. □

Suppose that I is an ideal in the ring $\mathbb{F}[X_1, \ldots, X_m]$. Let $P_1, \ldots, P_n$ be in the zero set of I with coordinates in $\mathbb{F}$. So $f(P_j) = 0$ for all $f \in I$ and all $j = 1, \ldots, n$. Then the evaluation map induces a well-defined linear map

$$ev_{\mathcal{P}} : \mathbb{F}[X_1, \ldots, X_m]/I \longrightarrow \mathbb{F}^n,$$

which is also a surjective morphism of $\mathbb{F}$-algebras.

Remark 4.4 In many papers *one point codes* are considered, that is to say codes of the form $C(D, mQ)$ or $C^*(D, mP)$, where Q is a rational point which is distinct from all $P_1, \ldots, P_n$ and m an integer. These codes are special cases of the construction in this section. Let R be the ring of $f \in \mathbb{F}_q(\mathcal{X})$ that have poles possibly in Q and nowhere else. Let $\rho(f) = -v_Q(f)$ as in Example 3.8. Let ρ_i be the ith nongap, see Definition 2.60. Then $E_l = C(D, \rho_l P)$ and $C_l = C^*(D, \rho_l P)$.

Example 4.5 In this example we discuss the question of the surjectivity of the evaluation map for curves with an affine equation

$$X^aY^c + uY^{b+c} + G = 0$$

over $\mathbb{F}_q$ with $u \in \mathbb{F}_q^*$, $G \in \mathbb{F}_q[X, Y]$, $\deg_X(G) = d < a$, $\deg(G) < b + c$ and $\gcd(a, b) = 1$, as treated in Example 3.16 and Proposition 3.17. Let $S = \mathbb{F}_q[X, Y]/(X^aY^c + uY^{b+c} + G)$. Then the map $ev_{\mathcal{P}} : S \to \mathbb{F}_q^n$ is surjective by Lemma 4.3. But S has no weight function if $c > 0$, as we have seen in Example 3.16. Let R be the subspace of S generated by the elements $x^\alpha y^\beta$ such that $\alpha < a$ and $c\alpha \leq (a - d)\beta$. Then R is a sub- $\mathbb{F}_q$-algebra of S with a weight function ρ such that $\rho(x^\alpha y^\beta) = \alpha b + \beta a$ by Proposition 3.17. Let φ

be the restriction of $ev_{\mathcal{P}}$ to R. Then φ is a morphism of $\mathbb{F}_q$-algebras, but it is not always surjective.

Take for instance the curve with equation $X^3Y + Y^3 + X^2 - 1 = 0$ over $\mathbb{F}_q$ of odd characteristic. This is of the above form with $a = 3$, $b = 2$, $c = 1$, $d = 2$ and $u = 1$, $G = X^2 - 1$. Let $\mathcal{P} = \{P_1, \dots, P_n\}$ be a collection of n distinct $\mathbb{F}_q$-rational points with $P_1 = (1, 0)$ and $P_2 = (-1, 0)$. The first two coordinates of $\varphi(f)$ are the same for all $f \in R$. Therefore the first two columns of the parity check matrix $(\varphi(f_i)_j | 1 \leq i \leq l, 1 \leq j \leq n)$ of C_l are the same. Hence $d(C_l) = 2$ for all l.

One can remedy this in two ways.

(1) The easiest way is to restrict the set $\mathcal{P}$. Assume that there is at most one point in $\mathcal{P}$ that lies on the line with equation $Y = 0$. One can show that this suffices to prove that φ is surjective.

(2) The second way needs some theory. We will explain it by means of an example. The $\mathbb{F}_q$-algebra R is again an affine algebra of a curve, since $R = \mathbb{F}_q[U, V]/(U^3 + V^5 + U^2 - V^2)$, where $U = XY$ and $V = Y$. This is seen by multiplying the original equation by Y^2. The original curve has the points $(1 : 0 : 0)$ and $(0 : 1 : 0)$ at infinity. The new curve has only $(1 : 0 : 0)$ at infinity. The map $(x : y : z) \mapsto (xy : yz : z^2)$ defines a morphism form the old curve to the new one. The three points $(1 : 0 : 1)$, $(-1 : 0 : 1)$ and $(1 : 0 : 0)$ are mapped to $(0 : 0 : 1)$. That is why the origin is a singularity of the second curve. One can extend the $\mathbb{F}_q$-algebra R to $\tilde{R}$ which still has a weight function. $\tilde{R}$ is the *normalisation* of R and consists of all fractions $f = a/b$ with $a, b \in R$ and $b \neq 0$, that satisfy an equation of the form $f^n + r_1 f^{n-1} + \cdots + r_{n-1} f + r_n = 0$, where $r_1, \dots, r_{n-1}, r_n \in R$. In the example above one can show that the subspace $\tilde{R}$ of S generated by R and $x + x^2y$ is the normalization. Let $f = x + xy^2$ and $r_2 = xy^4 + y^3 - xy - 1$. Then $f^2 + r_2 = 0$ and $r_2 \in R$. So indeed $f \in \tilde{R}$. The normalization $\tilde{R}$ is an $\mathbb{F}_q$-agebra with a weight function that extends the one on R. Let $\tilde{\varphi}$ be the restriction of $ev_{\mathcal{P}}$ to $\tilde{R}$. Then $\tilde{\varphi}$ is surjective. We will not show how to obtain $\tilde{R}$ in general.

In the setting of this section, the codes are very general and nothing specific can be said about the minimum distance of the codes E_l and C_l. This and the next section will show that certain order and weight functions on the affine ring R give a bound on the minimum distance which is in many cases the actual minimum distance. Furthermore Section 6 will show how to correct errors up to half the bound for C_l.

4.2 The order bound on the dual minimum distance

We repeat the main definitions. Let R be an $\mathbb{F}_q$-algebra with an order function ρ. Let $\{f_i \mid i \in \mathbb{N}\}$ be a basis of R over $\mathbb{F}_q$ such that $\rho(f_i) < \rho(f_{i+1})$ for all $i \in \mathbb{N}$. Let $\varphi : R \to \mathbb{F}_q^n$ be a surjective morphism of $\mathbb{F}_q$-algebra's. Let L_l be the vector space with $f_1, \ldots, f_l$ as a basis. The number $l(i,j)$ was defined as the smallest positive integer l such that $f_i f_j \in L_l$. The function $l(i,j)$ is strictly increasing in both arguments. Let $\mathbf{h}_i = \varphi(f_i)$. Let $E_l = \varphi(L_l)$ and C_l its dual. There exists a positive integer N such that $E_l = \mathbb{F}_q^n$ for all $l > N$. So $C_l = 0$ for all $l > N$. Let $\mathbf{H}$ be the $N \times n$ matrix with $\mathbf{h}_i$ as its i-th row for $1 \leq i \leq N$.

Definition 4.6 Let $\mathbf{y} \in \mathbb{F}_q^n$. Consider the *syndromes*

$$s_i(\mathbf{y}) = \mathbf{y} \cdot \mathbf{h}_i \quad \text{and} \quad s_{ij}(\mathbf{y}) = \mathbf{y} \cdot (\mathbf{h}_i * \mathbf{h}_j).$$

Then $\mathbf{S}(\mathbf{y}) = (s_{ij}(\mathbf{y}) \mid 1 \leq i,j \leq N)$ is the *matrix of syndromes* of $\mathbf{y}$.

Lemma 4.7 *Let $\mathbf{y} \in \mathbb{F}_q^n$. Let $\mathbf{D}(\mathbf{y})$ be the diagonal matrix with $\mathbf{y}$ on the diagonal. Then*

$$\mathbf{S}(\mathbf{y}) = \mathbf{H}\mathbf{D}(\mathbf{y})\mathbf{H}^T,$$

and

$$rank(\mathbf{S}(\mathbf{y})) = wt(\mathbf{y}).$$

Proof. The matrix of syndromes $\mathbf{S}(\mathbf{y})$ is equal to $\mathbf{H}\mathbf{D}(\mathbf{y})\mathbf{H}^T$, since

$$s_{ij}(\mathbf{y}) = \mathbf{y} \cdot (\mathbf{h}_i * \mathbf{h}_j) = \sum_l y_l h_{il} h_{jl},$$

where h_{il} is the l-th entry of $\mathbf{h}_i$. The rank of the diagonal matrix $\mathbf{D}(\mathbf{y})$ is equal to the number of nonzero entries of $\mathbf{y}$, which is $wt(\mathbf{y})$. The rows of $\mathbf{H}$ generate $\mathbb{F}_q^n$, since $E_N = \mathbb{F}_q^n$. Hence the matrices $\mathbf{H}$ and $\mathbf{H}^T$ both have full rank n. Therefore $\text{rank}(\mathbf{S}(\mathbf{y})) = \text{rank}(\mathbf{D}(\mathbf{y})) = wt(\mathbf{y})$. □

Definition 4.8 Let $l \in \mathbb{N}_0$. Define

$$N_l = \{\ (i,j) \in \mathbb{N}^2 \ \mid \ l(i,j) = l+1\ \}.$$

Let ν_l be the number of elements of N_l.

Lemma 4.9
(1) *If* $\mathbf{y} \in C_l$ *and* $l(i,j) \leq l$, *then* $s_{ij}(\mathbf{y}) = 0$.
(2) *If* $\mathbf{y} \in C_l \setminus C_{l+1}$ *and* $l(i,j) = l+1$, *then* $s_{ij}(\mathbf{y}) \neq 0$.

Proof.

(1) Let $\mathbf{y} \in C_l$. If $l(i,j) \leq l$, then $f_i f_j \in L_l$. So $\mathbf{h}_i * \mathbf{h}_j = \varphi(f_i f_j)$ is an element of $\varphi(L_l)$, which is the dual of C_l. So $s_{ij}(\mathbf{y}) = \mathbf{y} \cdot (\mathbf{h}_i * \mathbf{h}_j) = 0$.

(2) Let $\mathbf{y} \in C_l \setminus C_{l+1}$. If $l(i,j) = l+1$, then $f_i f_j \in L_{l+1} \setminus L_l$. So $f_i f_j \equiv \mu f_{l+1}$ modulo L_l for some nonzero $\mu \in \mathbb{F}_q$. Hence $\mathbf{h}_i * \mathbf{h}_j \equiv \mu \mathbf{h}_{l+1}$ modulo $\varphi(L_l)$. Now $\mathbf{y} \notin C_{l+1}$, so $s_{l+1}(\mathbf{y}) \neq 0$. Therefore $s_{ij}(\mathbf{y}) \neq 0$. □

Lemma 4.10 *If* $t = \nu_l$ *and* $(i_1, j_1), \ldots, (i_t, j_t)$ *is an enumeration of the elements of* N_l *in increasing order with respect to the lexicographic order on* $\mathbb{N}^2$, *then* $i_1 < \cdots < i_t$ *and* $j_t < \cdots < j_1$. *If moreover* $\mathbf{y} \in C_l \setminus C_{l+1}$, *then*

$$s_{i_u j_v}(\mathbf{y}) = \begin{cases} 0 & \textit{if} \quad u < v, \\ \textit{not zero} & \textit{if} \quad u = v. \end{cases}$$

Proof. The sequence $(i_1, j_1), \ldots, (i_t, j_t)$ is ordered in such a way that $i_1 \leq \ldots \leq i_t$ and $j_u < j_{u+1}$ if $i_u = i_{u+1}$. If $i_u = i_{u+1}$, then $j_u < j_{u+1}$, and therefore

$$l+1 = l(i_u, j_u) < l(i_u, j_{u+1}) = l(i_{u+1}, j_{u+1}) = l+1,$$

which is a contradiction. So the sequence $i_1, \ldots, i_t$ is strictly increasing. A similar argument shows that $j_{u+1} < j_u$ for all $u < t$.

Let $\mathbf{y} \in C_l$. If $u < v$, then $l(i_u, j_v) < l(i_v, j_v) = l+1$. Lemma 4.9 implies that $s_{i_u j_v}(\mathbf{y}) = 0$.

Moreover, let $\mathbf{y} \notin C_{l+1}$. If $u = v$, then $l(i_u, j_v) = l+1$. Lemma 4.9 implies that $s_{i_u j_v}(\mathbf{y}) \neq 0$. □

Proposition 4.11 *If* $\mathbf{y} \in C_l \setminus C_{l+1}$, *then* $wt(\mathbf{y}) \geq \nu_l$.

Proof. This follows from Lemmas 4.7 and 4.10. □

Definition 4.12

$$d(l) = \min\{\nu_m \mid m \geq l\},$$

$$d_\varphi(l) = \min\{\nu_m \mid m \geq l, C_m \neq C_{m+1}\},$$

The numbers $d(l)$ and $d_\varphi(l)$ will be called the *order* bound. If R is an affine algebra of the form $\mathbb{F}_q[X_1, \ldots, X_m]/I$ and φ is the evaluation map $ev_\mathcal{P}$ of the set $\mathcal{P}$ of n points in $\mathbb{F}_q^m$, then we denote d_φ by $d_\mathcal{P}$.

Theorem 4.13 *The numbers $d(l)$ and $d_\varphi(l)$ are lower bounds for the minimum distance of C_l:*

$$d(C_l) \geq d_\varphi(l) \geq d(l).$$

Proof. The theorem is a direct consequence of Definition 4.12 and Proposition 4.11. □

Remark 4.14 The set N_l and the numbers ν_l and $d(l)$ depend only on the order function ρ and neither on the choice of the basis $\{f_i \mid i \in \mathbb{N}\}$ nor on the choice of the set of points. The number $d_\mathcal{P}$ depends on the order function and the choice of the set of points, but not on the choice of the basis.

If $\mathcal{P} \subseteq \mathcal{P}'$, then $d_\mathcal{P} \geq d_{\mathcal{P}'}$.

Example 4.15 Let $R = \mathbb{F}_q[X]$ and let ρ, with $\rho(f) = \deg(f)$, be the order function of Example 3.7. Let $f_i = X^{i-1}$. For a primitive element α of $\mathbb{F}_q$ and $n = q-1$, let $\varphi : R \rightarrow \mathbb{F}_q^n$ be defined by $\varphi(f) = (f(\alpha^0), f(\alpha^1), \ldots, f(\alpha^{n-1}))$. Then $C_l = \{\mathbf{c} \in \mathbb{F}_q^n | \mathbf{c} \cdot \varphi(f_i) = 0, 1 \leq i \leq l\}$ then C_l is a cyclic code with defining set $\alpha^0, \alpha^1, \ldots, \alpha^{l-1}$. The order bound gives $d(l) = l+1$, from which the BCH bound may be derived.

Example 4.16 This is a continuation of Example 3.21 with the Klein quartic. The table gives a list of the functions f_l, their weights ρ_l, the numbers ν_l and the bound $d(l)$ from Theorem 4.13. We have that

$$N_l = \{(i,j) | \rho_i + \rho_j = \rho_{l+1}\},$$

since ρ is a weight function. It is easy to see that $d(l) = \nu_l = l-2$ for all $l \geq 6$.

l	1	2	3	4	5	6	7	8	9
f_l	1	y	xy	y^2	x^2y	xy^2	y^3	x^2y^2	xy^3
ρ_l	0	3	5	6	7	8	9	10	11
ν_l	2	2	3	2	4	4	5	6	7
$d(l)$	2	2	2	2	4	4	5	6	7

As a polynomial in Y, $X^3Y + Y^3 + X$, has three distinct zeros for every value of X in $\mathbb{F}_8^*$. The origin is the only point of the curve on the line $X = 0$ or $Y = 0$. Hence the affine curve has $21 + 1 = 22$ rational points. If we consider the evaluation codes, then for any set $\mathcal{P}$ of rational points the functions y^8 and y evaluate to the same vector. This gives $C_{21} = C_{22}$. If the set $\mathcal{P}$ consists of all the 21 rational points with nonzero coordinates, then the function y^7 evaluates to the same vector as the function 1, so furthermore $C_{18} = C_{19}$.

Let $l \geq 3$. Let $H_l = (\varphi(f_i)_j | 1 \leq i \leq l, 1 \leq j \leq n)$ be the parity check matrix of C_l. Then the first three entries of the j-th column of H_l are equal to 1, y_j and $x_j y_j$, where $P_j = (x_j, y_j)$. There is at most one j with $y_j = 0$. So any two columns of H_l are independent. Hence the minimum distance of the codes C_3 and C_4 is at least 3, and in fact it is equal to 3 for both codes. This is an example where the minimum distance of C_l is strictly larger than $d_{\mathcal{P}}(l)$ and $d(l)$.

Example 4.17 This is a continuation of Example 3.19 with the Hermitian curve over $\mathbb{F}_{16}$. Consider the table with a list of f_l, ρ_l, ν_l and $d(l)$.

l	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
f_l	1	x	y	x^2	xy	y^2	x^3	x^2y	xy^2	y^3	x^4	x^3y	x^2y^2	xy^3	y^4	x^4y
ρ_l	0	4	5	8	9	10	12	13	14	15	16	17	18	19	20	21
ν_l	2	2	3	4	3	4	6	6	4	5	8	9	8	9	10	12
$d(l)$	2	2	3	3	3	4	4	4	4	5	8	8	8	9	10	12

One can verify that $d(l) = \nu_l = l - 5$ for all $l > 16$.

Example 4.18 *Reed-Muller codes.* Let $R = \mathbb{F}_q[X_1, \ldots, X_m]$. Let ρ be the order function associated with the graded lexicographic order on the monomials of R. Let f_i be the i-th monomial with respect to this order. Let $n = q^m$. Let $P_1, \ldots, P_n$ be an enumeration of the n points of $\mathbb{F}_q^m = \mathcal{P}$. Then $RM_q(r, m)$ is by definition the code obtained by evaluating all $f \in \mathbb{F}_q[X_1, \ldots, X_m]$ of degree at most r at all points of $\mathcal{P}$. If $f_l = X_1^r$, then $f_{l+1} = X_m^{r+1}$ and $\{f_i \mid i \leq l\}$ is the set of all monomials of degree at most r. So $RM_q(r, m) = ev_{\mathcal{P}}(L_l) = E_l$. The minimum distance of Reed-Muller codes is well-known. It is also a consequence of the theory developed above, as we will now demonstrate.

Lemma 4.19
(1) *If $f_{l+1} = X^\gamma$, then $\nu_l = \prod_{t=1}^m (\gamma_t + 1)$.*
(2)

$$d(l) = deg(f_l) + \left\{ \begin{array}{ll} 2 & \textit{if } f_l = X_1^r \textit{ for some } r, \\ 1 & \textit{otherwise.} \end{array} \right.$$

(3) *Let $f_l = X_1^r$. Write $r+1 = \nu(q-1)+\mu$ with $\nu, \mu \in \mathbb{N}_0$ such that $\mu < q-1$. Then $d_{\mathcal{P}}(l) = (\mu + 1)q^\nu$.*

Proof.

(1) If $f_i = X^\alpha$, $f_j = X^\beta$, then $f_l = X^{\alpha+\beta}$ for some l. So $l(i,j) = l$. Hence if $f_{l+1} = X^\gamma$, then ν_l is equal to the number of pairs (i,j) such that $f_i f_j = f_{l+1}$, which is equal to the number of l $\alpha \in \mathbb{N}_0^m$ such that $0 \leq \alpha_t \leq \gamma_t$ for all t, $1 \leq t \leq m$, which is $\prod_{t=1}^m (\gamma_t + 1)$.

(2) If $f_l = X_1^r$, then $f_{l+1} = X_m^{r+1}$. So $\nu_l = r + 2 = \deg(f_l) + 2$. Let $l' \geq l$ and $f_{l'+1} = X^\gamma$. Then

$$\nu_{l'} = \prod_{t=1}^m (\gamma_t + 1) \geq (\sum_{t=1}^m \gamma_t) + 1 = \deg(f_{l'+1}) + 1 \geq \deg(f_l) + 2.$$

So $d(l) = \deg(f_l) + 2$.
If f_l is not of the form X_1^r, then $f_{l_0+1} = X_1^r$ for some $l_0 \geq l$ and $r = \deg(f_l)$. So $\nu_{l_0} = r + 1$ and $\nu_{l'} \geq r + 1$ for all $l' \geq l$. Hence $d(l) = \deg(f_l) + 1$.

(3) If $f_{l'+1} = X^\gamma$, then the code $C_{l'}$ is not equal to $C_{l'+1}$ if and only if $0 \leq \gamma_t \leq q - 1$ for all t. So $d_{\mathcal{P}}(l)$ is equal to

$$\min\{ \ \prod_{t=1}^m (\gamma_t + 1) \mid \sum_{t=1}^m \gamma_t \geq r + 1 \text{ and } 0 \leq \gamma_t \leq q - 1 \text{ for all } t \ \},$$

if $f_l = X_1^r$. Consider the real valued function f defined by $f(\mathbf{x}) = \prod_{t=1}^m (x_t + 1)$ on the domain $\{\mathbf{x} \in \mathbb{R}^m \mid \sum_{i=1}^m x_i \geq r + 1 \text{ and } 0 \leq x_t \leq q - 1 \text{ for all } t \ \}$. The method of Lagrangemultipliers gives that the minimum of f is obtained at the corner $(0, \ldots, 0, \mu, q - 1, \ldots, q - 1)$, where the last ν cordinates are equal to $q - 1$. Hence $d_{\mathcal{P}}(l) = (\mu + 1)q^\nu$. □

Theorem 4.20 *Let r and m be positive integers such that $0 \leq r < (q-1)m$. Write $(q-1)m - r = \nu(q-1) + \mu$ with $\nu, \mu \in \mathbb{N}_0$ such that $\mu < q - 1$. Then the minimum distance of $RM_q(r,m)$ is equal to $(\mu + 1)q^\nu$.*

Proof. Let E_l be the Reed-Muller code $RM_q((q-1) - r - 1, m)$ and obtained as described in Example 4.18. Then C_l is the dual of this code if $f_l = X_1^{(q-1)-r-1}$.

The dual of $RM_q(r,m)$ is a RM code of order $m(q-1) - r - 1$. We give a sketch of the proof of this fact.

Notice that X_i^q and X_i evaluate to the same word under $ev_{\mathcal{P}}$. A polynomial is called *reduced* if the monomials X^α with a nonzero coefficient satisfy $\alpha_i < q$ for $i = 1, \ldots, m$. So for every polynomial F there exists a reduced polynomial F' such that $ev_{\mathcal{P}}(F) = ev_{\mathcal{P}}(F')$. This polynomial F' is unique. So the dimension of $RM_q(r,m)$ is equal to the number of reduced monomials

X^α such that $\deg(\alpha) \leq r$. Let $Q = \{0, 1, \dots, q-1\}$. Let $\mu = (q-1, \dots, q-1)$. By considering the map $X^\alpha \mapsto X^{\mu-\alpha}$ on Q^m one sees that

$$\dim RM_q(r, m) + \dim RM_q((q-1)m - r - 1, m) = q^m.$$

Let F and G be two reduced polynomials. Then

$$ev_{\mathcal{P}}(F) \cdot ev_{\mathcal{P}}(G) = 0 \text{ if } \deg(F) + \deg(G) < (q-1)m.$$

This is seen by considering monomials first and using the fact that $\sum_{x \in \mathbb{F}_q} x^i = 0$ for all $i < q-1$. Hence the two RM codes are orthogonal. By the above remark on the dimensions we have proved that the two RM codes are dual to each other.

Hence $C_l = RM_q(r, m)$. Now $(q-1)m - r = (q-1)\nu + \mu$. So the minimum distance of $RM_q(r, m)$ is at least $(\mu+1)q^\nu$ by Lemma 4.19. To show that this lower bound is tight, we consider the polynomial

$$F = \prod_{i=1}^{m-\nu-1} \left(X_i^{q-1} - 1\right) \prod_{j=\mu+1}^{q-1} \left(X_{m-\nu} - a_j\right),$$

where $\mathbb{F}_q = \{a_0, \dots, a_{q-1}\}$. Then $\deg(F) = (m-\nu-1)(q-1) + (q-1-\mu) = r$. So F evaluates to a codeword of $RM_q(r, m)$. Let $P = (x_1, \dots, x_m)$. Then $F(P) \neq 0$ if and only if $x_1 = \cdots = x_{m-\nu-1} = 0$ and $x_{m-\nu} = a_j$ for some $j \in \{0, \dots, \mu\}$. Hence $ev_{\mathcal{P}}(F)$ has weight $(\mu+1)q^\nu$. □

4.3 Improvements and generalizations

In this section we sketch, without proofs, possible improvements and generalizations of the theory.

Let R be an $\mathbb{F}_q$-algebra. Let $(f_i | i \in \mathbb{N})$, $(g_j | j \in \mathbb{N})$ and $(h_l | l \in \mathbb{N})$ be three sequences of independent elements in R. The vector space $L(l)$ has $h_1, \dots, h_l$ as basis. Assume that for all i, j there exists an l such that $f_i g_j \in L(l)$. The function $l(i, j)$ is defined as the smallest l such that $f_i g_j \in L(l)$. Assume that $l(i, j)$ is strictly increasing in both arguments. In the theory of the previous sections we have the special case with $f_i = g_i = h_i$, $(f_i | i \in \mathbb{N})$ is a basis of R and $\rho(f_i) < \rho(f_{i+1})$ for all i, where ρ is an order function on R. Let

$\varphi : R \longrightarrow \mathbb{F}_q^n$ be a surjective morphism of $\mathbb{F}_q$-algebras. Let $\mathbf{h}_l = \varphi(h_l)$. Let E_l be the code generated by $\mathbf{h}_1, \ldots, \mathbf{h}_l$. Let C_l be the dual of E_l. So

$$C_l = \{\mathbf{c} \in \mathbb{F}_q^n \mid \mathbf{c} \cdot \mathbf{h}_i = 0 \text{ for all } i \leq l\}.$$

The set N_l is defined by

$$N_l = \{\ (i,j) \in \mathbb{N}^2 \mid l(i,j) = l+1\ \}$$

as in Definition 4.8, and ν_l is the number of elements of N_l. Let

$$d(l) = \min\{\nu_m \mid m \geq l\},$$

$$d_\varphi(l) = \min\{\nu_m \mid m \geq l, C_m \neq C_{m+1}\}$$

as in Definition 4.12. Then $d_\varphi(l) \geq d(l)$ and they are lower bounds on the minimum distance of the code C_l.

Example 4.21 Consider the curve with equation $X^3Y + Y^3 + X^2 - 1 = 0$ as in Example 4.5. Let the sequence $(h_l | l \in \mathbb{N})$ list the elements $x^\alpha y^\beta$, $\alpha < 3$ such that $2\alpha + 3\beta$ is increasing. So $h_1 = 1$, $h_2 = x$, $h_3 = y$ and $h_{3k-e} = x^e y^{k-e}$ if $k \geq 2$ and $e = 0, 1, 2$. Let $(f_i | i \in \mathbb{N})$ be the sequence obtained from $(h_l | l \in \mathbb{N})$ by deleting x^2. So $f_1 = 1$, $f_2 = x$, $f_3 = y$, $f_4 = xy$, $f_5 = y^2$ and $f_{3k+e} = x^{2-e} y^{k+e-1}$ if $k \geq 2$ and $e = 0, 1, 2$. Let $g_i = f_i$. One can verify that the product $f_i g_j$ is a linear combination of the h_l, and that the function $l(i,j)$ is strictly increasing in both arguments. It is easy to see that $d(l) = \nu_l = l - 2$ for all $l \geq 8$.

l	1	2	3	4	5	6	7	8	9	10
f_l	1	x	y	xy	y^2	x^2y	xy^2	y^3	x^2y^2	xy^3
h_l	1	x	y	x^2	xy	y^2	x^2y	xy^2	y^3	x^2y^2
ν_l	2	2	1	4	3	4	6	6	7	8
$d(l)$	1	1	1	3	3	4	6	6	7	8

Let $\mathbb{F}_0$ be a subfield of $\mathbb{F}_q$. In the above situation we can define the code

$$C_l^0 = \{\mathbf{c} \in \mathbb{F}_0^n \mid \mathbf{c} \cdot \mathbf{h}_i = 0 \text{ for all } i \leq l\}.$$

Then C_l^0 is a subfield subcode of C_l. So the bounds $d(l)$ and $d_\varphi(l)$ hold also for the minimum distance of C_l^0. Define

$$d_\varphi^0(l) = \min\{\nu_m \mid m \geq l, C_m^0 \neq C_{m+1}^0\}.$$

Then $d^0_\varphi(l) \geq d_\varphi(l) \geq d(l)$ and they are lower bounds on the minimum distance of the code C^0_l. In this way one get bounds on the minimum distance of cyclic codes that improve the BCH bound. The most general bound is the so called *shift* bound. We will not define it here but refer to the Notes. All these bounds have the decomposition of the matrix of syndromes $S(\mathbf{y})$ in common, and patterns of zeros in this matrix give information on the nonzeros of $\mathbf{y}$.

It is possible to give a version of these ideas on the level of words in $\mathbb{F}_q^n$ directly without any reference to the $\mathbb{F}_q$-algebra R and the morphism φ.

Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be three bases of $\mathbb{F}_q^n$. Let $\bar{E}_l$ be the code generated by $\mathbf{c}_1, \ldots, \mathbf{c}_l$. Let $\bar{C}_l$ be the dual of the code $\bar{E}_l$. Let $\bar{l}(i,j)$ be the the smallest positive integer l such that $\mathbf{a}_i * \mathbf{b}_j \in \bar{E}_l$. The pair (i,j) is called *well-behaving* if $\bar{l}(i',j') < \bar{l}(i,j)$ for all i', j' such that $i' \leq i$, $j' \leq j$ and $(i',j') \neq (i,j)$. Let $l = 0, 1, \ldots, n-1$. Define

$$\bar{N}(l) = \{ \ (i,j) \mid \bar{l}(i,j) = l+1 \text{ and } (i,j) \text{ is well-behaving } \ \}.$$

Let $\bar{\nu}(l)$ be the number of elements of $\bar{N}(l)$. Let $\bar{\nu}(n) = n+1$. Define

$$\bar{d}(l) = \min\{ \ \bar{\nu}(m) \mid l \leq m \leq n \ \}.$$

Then $\bar{d}(l)$ is a lower bound on the minimum distance of $\bar{C}_l$.

Let the basis $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ be obtained by deleting succesively superfluous elements of the sequence $(\varphi(f_i) | i \in \mathbb{N})$, and let the bases $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be obtained similarly from $(\varphi(g_j) | j \in \mathbb{N})$ and $(\varphi(h_l) | l \in \mathbb{N})$, respectively. If the dimension of C_l is k, $r = n-k$ and $C_l \neq C_{l+1}$, then $\bar{C}_r = C_l$ and $\bar{d}(r) \geq d_\varphi(l)$.

Definition 4.22 Let d be a positive integer. Define

$$\tilde{C}(d) = \{\mathbf{c} \in \mathbb{F}_q^n \mid \mathbf{c} \cdot \mathbf{h}_{l+1} = 0 \text{ for all } l \in \mathbb{N}_0 \text{ such that } \nu_l < d\},$$

$$\tilde{C}_\varphi(d) = \{\mathbf{c} \in \mathbb{F}_q^n \mid \mathbf{c} \cdot \mathbf{h}_{l+1} = 0 \text{ for all } l \in \mathbb{N}_0 \text{ such that } \nu_l < d \text{ and } C_l \neq C_{l+1}\}.$$

Proposition 4.23 *The minmum distance of $\tilde{C}(d)$ and $\tilde{C}_\varphi(d)$ is at least d.*

Proof. The code $\tilde{C}(d)$ is contained in $\tilde{C}_\varphi(d)$. So it is enough to prove the claim for the latter. Let $\mathbf{y}$ be a nonzero codeword of $\tilde{C}_\varphi(d)$. If $d = 1$, then there is nothing to prove. Let $d > 1$. Then $\nu_0 = 1 < d$. So $\mathbf{y} \cdot \mathbf{h}_1 = 0$. The number N was defined in such a way that the elements $\mathbf{h}_1, \ldots, \mathbf{h}_N$ generate $\mathbb{F}_q^n$. The word $\mathbf{y}$ is not zero. So there exists a positive integer l such that $\mathbf{y} \cdot \mathbf{h}_{l+1} \neq 0$. Let l be the smallest positive integer such that $\mathbf{y} \cdot \mathbf{h}_{l+1} \neq 0$. Then $\mathbf{y} \in C_l \setminus C_{l+1}$. Therefore $wt(\mathbf{y}) \geq \nu_l$ by Proposition 4.11. If $\nu_l < d$, then $\mathbf{y} \cdot \mathbf{h}_{l+1} = 0$, since $\mathbf{y} \in \tilde{C}_\varphi(d)$ and $C_l \neq C_{l+1}$. This is a contradiction. Hence $wt(\mathbf{y}) \geq \nu_l \geq d$. □

Definition 4.24 Let d be a positive integer. Define

$$R(d) = \{l + 1 \mid l \in \mathbb{N}_0,\ \nu_l < d\},$$

$$R_\varphi(d) = \{l + 1 \mid l \in \mathbb{N}_0,\ \nu_l < d \text{ and } C_l \neq C_{l+1}\}.$$

Let $r(d)$ and $r_\varphi(d)$ be the number of elements of $R(d)$ and $R_\varphi(d)$, respectively.

Remark 4.25 The number $r(d)$ is the number of paritychecks that define $\tilde{C}(d)$ and depends only on the order function. These paritychecks might be dependent. So the redundancy of $\tilde{C}(d)$ is at most $r(d)$. Hence the dimension of $\tilde{C}(d)$ is at least $n - r(d)$. The number $r_\varphi(d)$ is the number of paritychecks that define $\tilde{C}_\varphi(d)$, and depends on the order function and the map φ. In this case these paritychecks are independent by definition. So the redundancy of $\tilde{C}_\varphi(d)$ is equal to $r_\varphi(d)$. Hence the dimension of $\tilde{C}_\varphi(d)$ is $n - r_\varphi(d)$.

The codes $\tilde{C}(d)$ and $\tilde{C}_\varphi(d)$ have the *super code property*, that is to say: if $d = d(l)$, then $C_l \subseteq \tilde{C}(d) \subseteq \tilde{C}_\varphi(d)$. So the minimum distance of the codes C_l, $\tilde{C}(d)$ and $\tilde{C}_\varphi(d)$ is at least d, but C_l might be smaller.

Example 4.26 *Hyperbolic codes.* Consider the situation as in Example 4.18 of the RM codes. So $n = q^m$ and $\mathcal{P} = \{P_1, \ldots, P_n\}$ is an enumertion of the points of the affine space $\mathbb{F}_q^m$. Furthermore $R = \mathbb{F}_q[X_1, \ldots, X_m]$ and φ is the evaluation map $ev_{\mathcal{P}}$. We have seen that that all the *RM* codes are of the form E_l and C_l, and that the order bound is equal to the minimum distance by Theorem 4.20. The bounds on the redundancies are given by

$$r(d) = \#\{\alpha \in \mathbb{N}_0^m \mid \textstyle\prod_{i=1}^m (\alpha_i + 1) < d\},$$

$$r_\varphi(d) = \#\{\alpha \in Q^m \mid \textstyle\prod_{i=1}^m (\alpha_i + 1) < d\},$$

where $Q = \{0, 1, \ldots, q - 1\}$, by Lemma 4.19. The codes $\tilde{C}(d)$ and $\tilde{C}_\varphi(d)$ are called *hyperbolic*, since in case $m = 2$ the bound $r(d)$ is the number of integer lattice points $(x, y) \in \mathbb{N}_0^m$ under the hyperbola $(x + 1)(y + 1) = d$.

4.4 Notes

The order bound uses the fact that the matrix of syndromes is a triple product of matrices which stems from the study of decoding algorithms as will be seen in Section 6. The order bound is also called the *Feng-Rao* bound. It can be seen as a generalisation of the *shift* bound for cyclic codes [63, 74].

The generalized RM codes are treated in [6, 15, 52] and [94].

The idea to define $\tilde{C}(d)$ by deleting parity checks in the definition of C_l and still keeping the same bound on the minimum distance is from [25], where they are called improved geometric Goppa codes. From an algebraic geometric point of view these codes are defined with the help of incomplete linear systems.

Hyperbolic codes are called Hyperbolic Cascaded Reed Solomon codes in [82].

The idea to generalize the construction of Goppa from curves to varieties of arbitrary dimensions is from [66], but until the order bound, not much could be said about the parameters of these codes if the dimension of the variety is greater than one.

5 Weight functions and semigroups

The order bound is investigated in detail when ρ is a weight function in terms of its associated semigroup, in particular if the semigroup is generated by two generators, and more generally, for semigroups which are called telescopic. The minimum distance of Hermitian codes is determined.

5.1 Semigroups and the minimum distance

Suppose that ρ is a weight function. Condition $(O.5)$ implies that the subset $\Lambda = \{ \rho(f) \mid f \in R, f \neq 0 \}$ of the nonnegative integers $\mathbb{N}_0$ has the property that, $0 \in \Lambda$, and $x + y \in \Lambda$ for all $x, y \in \Lambda$.

Definition 5.1 A subset Λ of $\mathbb{N}_0$ is called a *(numerical) semigroup* if $0 \in \Lambda$ and for all $x, y \in \Lambda$ also the sum $x + y \in \Lambda$.

Elements of $\mathbb{N}_0 \setminus \Lambda$ are called *gaps* of Λ and elements of Λ are called *nongaps* of Λ. If all elements of Λ are divisible by an integer $d > 1$, then there are infinitely many gaps. The *number of gaps* is denoted by $g = g(\Lambda)$.

If $g < \infty$, then there exists an $n \in \Lambda$ such that if $x \in \mathbb{N}_0$ and $x \geq n$, then $x \in \Lambda$. The *conductor* of Λ is the smallest $n \in \Lambda$ such that $\{ \ x \in \mathbb{N}_0 \mid x \geq n \ \}$ is contained in Λ, denoted by $c = c(\Lambda)$. So $c - 1$ is the *largest gap* of Λ if $g > 0$.

Example 5.2 If ρ is a weight function, then $\Lambda = \{ \ \rho(f) \mid f \in R, f \neq 0 \ \}$ is the semigroup of ρ.

In particular, if $\rho = -v_P$ of Example 3.8, then Λ is the Weierstrass semigroup of P, see Definition 2.15.

Remark 5.3 Let Λ be a semigroup with g gaps and conductor c.
(1) $g = 0$ if and ony if $c = 0$.
(2) Let $g > 0$. Then $c \geq g + 1$, and $\Lambda = \{ \ x \in \mathbb{N}_0 \mid x \geq g + 1 \ \} \cup \{0\}$ if and only if $c = g + 1$.
(3) There is exactly one gap if and only if 1 is the only gap.
(4) If 2 is a nongap, then $\{1, 3, \ldots, 2g - 1\}$ is the set of gaps. So $c = 2g$.

Example 5.4 Let $\Lambda = \{0, 4, 5, 8, 9, 10\} \cup \{ \ x \in \mathbb{N}_0 \mid x \geq 12 \ \}$. Then Λ is the semigroup of the weight function on the Hermitian curve over $\mathbb{F}_{16}$ of Example 3.19. The gaps are $1, 2, 3$, $6, 7$ and 11. So the number of gaps is $g = 6$, the conductor is $c = 12$ and the largest gap is 11.

Definition 5.5 The elements of a semigroup Λ will be enumerated by the sequence $(\rho_l | l \in \mathbb{N})$ such that $\rho_l < \rho_{l+1}$ for all l. The number of gaps smaller than ρ_l will be denoted by $g(l)$.

Lemma 5.6 *Let Λ be a semigroup with finitely many gaps.*
(1) *If $l \in \mathbb{N}$, then $g(l) = \rho_l - l + 1$.*
(2) *If $l \in \mathbb{N}$, then $\rho_l \leq l + g - 1$ and equality holds if and only if $\rho_l \geq c$.*
(3) *If $l > c - g$, then $\rho_l = l + g - 1$.*
(4) *If $l \leq c - g$, then $\rho_l < c - 1$.*

Proof.

(1) The nongap ρ_l is the (ρ_l+1)-st element of $\mathbb{N}_0$. So ρ_l is the $(\rho_l+1-g(l))$-th element of the semigroup Λ. Hence $l = \rho_l + 1 - g(l)$.

(2) Now $g(l) \leq g$ and $g(l) = g$ if and only if $\rho_l \geq c$.

(3) The conductor c is the $(c + 1)$-st element of $\mathbb{N}_0$. All gaps are strictly smaller than c. So c is the $(c + 1 - g)$-th element of Λ. Hence $c = \rho_{c+1-g}$. Let $l > c - g$. Then $\rho_l \geq \rho_{c-g+1} = c$. Therefore $\rho_l = l + g - 1$ by (2).

(4) Let $l \leq c - g$. Then $\rho_l \leq l + g - 1 \leq c - 1$. But $c - 1$ is a gap or negative. Therefore $\rho_l < c - 1$. □

In the previous lemma we used only the fact that $\{ n \in \mathbb{N}_0 \mid n \geq c \}$ is contained in Λ and $c - 1 \notin \Lambda$. In the following proposition we use the property that a semigroup is closed under addition.

Proposition 5.7 *Suppose that the number of gaps is finite. Then*

$$c \leq 2g.$$

And $c = 2g$ if and only if for any nonnegative integer s, if s is a gap, then $c - 1 - s$ is a nongap.

Proof. Consider a pair of nonnegative integers (s, t) with $s + t = c - 1$. At least one of these two numbers has to be a gap, since $c - 1$ is a gap and the sum of two nongaps is a nongap. But there are c such pairs, giving the required inequality.

Equality holds if and only if for any pair of nonnegative integers (s, t) with $s + t = c - 1$ exactly one of these two numbers is a nongap and the other is a gap. □

Definition 5.8 A semigroup is called *symmetric* if $c = 2g$.

Example 5.9 The semigroup of the weight function of the Klein curve of Example 3.21 has three gaps: $1, 2$ and 4. The largest gap is 4 and the conductor is 5. So this semigroup is not symmetric.

Definition 5.10 Let $A = \{a_1, \ldots, a_k\}$ be a subset of a semigroup Λ. If for any element $s \in \Lambda$ there exist $x_1, \ldots, x_k \in \mathbb{N}_0$ such that $s = \sum_{i=1}^{k} x_i a_i$, the semigroup Λ is said to be *generated* by A and written $\Lambda = \langle A \rangle$. A set A of generators of Λ is *minimal* if Λ is not generated by a proper subset of A.

We state without proof the following facts. Every semigroup has a finite set of generators. Every set of generators contains a minimal set of generators and a minimal set of generators is unique.

Proposition 5.11 *Let $a, b \in \mathbb{N}$ such that $gcd(a, b) = 1$. The semigroup generated by a and b is symmetric, has $ab - a - b$ as largest gap, $(a-1)(b-1)$ as conductor and the number of gaps is equal to $(a - 1)(b - 1)/2$.*

Proof. Every integer m has a unique representation $m = xb + ya$, where x and y are integers such that $0 \leq y < b$, since $\gcd(a, b) = 1$. Hence every gap m has a unique representation $m = xb + ya$ such that $0 \leq y < b$ and $x < 0$, and every nongap m has a unique representation $m = xb + ya$ such that $0 \leq y < b$ and $x \geq 0$.

Let c be the conductor of the semigroup $\Lambda = \langle a, b \rangle$. First the largest gap $c - 1$ is computed. The numbers $ya \in \Lambda$, $y = 0, 1, \ldots, b - 1$ form a complete set of representatives of the cosets modulo b, and $ya - b$ is the largest element in the coset of ya without a representation with nonnegative integer coefficients. Hence

$$(b - 1)a - b$$

is the largest gap, which is equal to $c - 1$. So $c = (a - 1)(b - 1)$. To see that $\langle a, b \rangle$ is symmetric, assume that s and t are both gaps and $s + t = c - 1$. Since s and t can be written as

$$s = x_1 b + y_1 a, \quad t = x_2 b + y_2 a, \quad 0 \leq y_1, y_2 < b \ \text{ and } \ x_1, x_2 < 0,$$

we get $c - 1 = ab - a - b = (x_1 + x_2)b + (y_1 + y_2)a$. So

$$(-x_1 - x_2 - 1)b = (y_1 + y_2 - b + 1)a,$$

where $0 \leq y_1 + y_2 \leq 2b - 2$ and $x_1 + x_2 \leq -2$. The lefthand side of the last equation is strictly positive and the righthand side strictly smaller than ab, giving a contradiction, because $\gcd(a, b) = 1$. Hence Λ is symmetric and $c = 2g$ by Proposition 5.7, where g is the number of gaps. So $g = (a-1)(b-1)/2$. □

Corollary 5.12 *A semigroup has a finite number of gaps if and only if the greatest common divisor of its elements is* 1.

Proof. Suppose that the greatest common divisor of the elements of a semigroup Λ is 1. Then there exist $a, b \in \Lambda$ such that $gcd(a, b) = 1$. The number of gaps of $\langle a, b \rangle$ is finite by Proposition 5.11, and Λ contains $\langle a, b \rangle$. So the number of gaps of Λ is finite.

The converse is clear. □

Example 5.13 The semigroup of Example 5.4 is generated by 4 and 5. This semigroup has $6 = (4 - 1)(5 - 1)/2$ gaps and is indeed symmetric.

Example 5.14 The semigroup of the weight function of plane curves with defining equation $X^aY^c + Y^{b+c} + G = 0$ as treated in Proposition 3.17, is equal to

$$\langle a, b\rangle \setminus \{ \ \alpha b + \beta a \mid \alpha, \beta \in \mathbb{N}_0, \ \alpha < a, c\alpha > (a-d)\beta \ \}.$$

Hence $g = (a-1)(b-1)/2 + \Delta$, where Δ is equal to the number of elements of $\{(\alpha, \beta) \in \mathbb{N}_0^2 | \alpha < a, c\alpha > (a-d)\beta\}$.

For a subset B of $\mathbb{N}_0$ and $a \in \mathbb{N}_0$ we denote the set $\{ \ a+b \mid b \in B \ \}$ by $a+B$.

Lemma 5.15 *Let Λ be a semigroup with finitely many gaps. Let $s \in \Lambda$. Then the number of elements of $\Lambda \setminus (s+\Lambda)$ is equal to s.*

Proof. Let c be the conductor of Λ. Let $T = \{ \ t \in \mathbb{N}_0 \mid t \geq s+c \ \}$. Then T is contained in Λ and in $s+\Lambda$. Let $U = \{ \ u \in \Lambda \mid u < s+c \ \}$. Then the number of elements of U is equal to $s+c-g$, and Λ is the disjoint union of T and U. Let $V = \{ \ v \in s+\Lambda \mid s \leq v < s+c \ \}$. Then the number of elements of V is equal to $c-g$, and $s+\Lambda$ is the disjoint union of V and T. Furthermore $V \subseteq U$, since $s \in \Lambda$ and Λ is a semigroup. Hence

$$\#(\Lambda \setminus (s+\Lambda)) = \#U - \#V = (s+c-g) - (c-g) = s.$$

□

Lemma 5.16 *Let f be a nonzero element of an $\mathbb{F}_q$-algebra R with a weight function ρ. Then*

$$dim(R/(f)) = \rho(f).$$

Proof. Let Λ be the semigroup of the weight function ρ. Let $s = \rho(f)$. Let $(\rho_i \mid i \in \mathbb{N})$ be the sequence of the elements of Λ in increasing order. The image under ρ of the set of nonzero elements of the ideal (f) is equal to $s+\Lambda$. So for every $\rho_i \in \Lambda$ there exists an $f_i \in R$ such that $\rho(f_i) = \rho_i$. If moreover $\rho_i \in s+\Lambda$, then we may choose $f_i \in (f)$. The sets $\{ \ f_i \mid i \in \mathbb{N} \ \}$ and $\{ \ f_i \mid i \in \mathbb{N}, \ \rho_i \in s+\Lambda \ \}$ are bases of the algebra R and the ideal (f), respectively, by the same argument as 3.12. Hence the classes of f_i modulo (f) with $i \in \mathbb{N}$ and $\rho_i \in \Lambda \setminus (s+\Lambda)$ form a basis of $R/(f)$. So the dimension of $R/(f)$ is equal to the number of elements of $\Lambda \setminus (s+\Lambda)$, which is $\rho(f)$ by Lemma 5.15. □

Lemma 5.17 *Let R be an affine algebra with a weight function ρ and an evaluation map $ev_{\mathcal{P}}$. Let f be a nonzero element of R. Then the number of zeros of f is at most $\rho(f)$.*

Proof. Let $\mathcal{Q}$ be the set of zeros of f and let $t = |\mathcal{Q}|$. The map $ev_{\mathcal{Q}} : R \rightarrow \mathbb{F}_q^t$ is linear and surjective by Lemma 4.3. Furthermore $g(Q) = 0$ for all $Q \in \mathcal{Q}$ and $g \in (f)$. This induces a well-defined map $ev_{\mathcal{Q}} : R/(f) \rightarrow \mathbb{F}_q^t$ which is linear and surjective. So the number of zeros of f is at most the dimension of $R/(f)$ which is equal to $\rho(f)$ by Lemma 5.16. □

Suppose that we have a weight function ρ on $R = \mathbb{F}_q[X_1, \ldots, X_m]/I$. Let $(\rho_i | i \in \mathbb{N})$ be the enumeration of the elements of the semigroup of ρ in increasing order. Let $\mathcal{P}$ consist of n distinct points of $\mathbb{F}_q^m$ in the zero set of I, and let $ev_{\mathcal{P}} : R \rightarrow \mathbb{F}_q^n$ be the corresponding evaluation map. The evaluation code E_l is defined as in Section 4.1. So $E_l = \{ \ ev_{\mathcal{P}}(f) \mid f \in R, \rho(f) \leq \rho_l \ \}$.

Theorem 5.18 *The minimum distance of E_l is at least $n - \rho_l$. If $\rho_l < n$, then $dim(E_l) = l$.*

Proof. Let $\mathbf{c}$ be a nonzero element of E_l. Then there exists a nonzero element $f \in R$ such that $\rho(f) \leq \rho_l$ and $\mathbf{c} = ev_{\mathcal{P}}(f)$. So $c_i = f(P_i)$ for all i. The number of zeros of f is at most ρ_l, by Lemma 5.17. So $wt(\mathbf{c}) \geq n - \rho_l$.

Suppose moreover that $\rho_l < n$. E_l is the image under the evaluation map of the vector space L_l of dimension l. If $f \in L_l$ and $ev_{\mathcal{P}}(f) = 0$, then f has at least n zeros. Hence $f = 0$ by Lemma 5.17, since $\rho_l < n$. So the map $ev_{\mathcal{P}} : L_l \rightarrow E_l$ is a linear isomorphism, hence $\dim(E_l) = l$. □

Corollary 5.19 *Let ρ be a weight function with g gaps. If $\rho_k < n$, then E_k is an $[n, k, d]$ code such that $k + d \geq n + 1 - g$.*

Proof. This follows from Theorem 5.18 and the fact that $\rho_k \leq k + g - 1$ as shown in Lemma 5.6. □

Example 5.20 This is a continuation of Example 5.14. In the special case that $a = b + 1$ and $c = 0$, one can compare the code C of Section 2.3 with $m = a$ and the code E_k of this section. The number of gaps is $g = \binom{m-1}{2}$ by Proposition 5.11. If $\rho_k = lm$, then $C = E_k$ and Proposition 2.27 and Corollary 5.19 give the same parameters of the codes C and E_k, respectively.

Remark 5.21 If ρ is an order function but not a weight function, then in general $R/(f)$ is not finite-dimensional and there is not a straightforward bound on the minimum distance for E_l.

5.2 Semigroups and the dual minimum distance

Let ρ be a weight function on the $\mathbb{F}$-algebra R. It is assumed that the greatest common divisor of the weights $\rho(f)$, $0 \neq f \in R$, is 1. So g, the number of gaps of the corresponding semigroup Λ, is finite. Let $(\rho_i \mid i \in \mathbb{N})$ be the sequence of nongaps of the weight function ρ such that $\rho_i < \rho_{i+1}$ for all i. The number of gaps smaller than ρ_l is denoted by $g(l)$. The conductor of Λ is denoted by c.

Recall from Proposition 3.12 that for a weight function ρ the function $l(i,j)$ is determined by

$$\rho_{l(i,j)} = \rho_i + \rho_j.$$

Hence the set N_l from Definition 4.8 can be redefined by

$$N_l = \{ \ (i,j) \in \mathbb{N}^2 \mid \rho_i + \rho_j = \rho_{l+1} \ \}.$$

The number of elements of N_l is denoted by ν_l. In Definition 4.12 the order bound $d(l)$ was defined by

$$d(l) = \min\{ \ \nu_m \mid m \geq l \ \}.$$

Definition 5.22 The *Goppa bound* on the minimum distance of C_l is denoted by $d_G(l)$ and is defined by $d_G(l) = l + 1 - g$.

Example 5.23 If $\Lambda = \mathbb{N}_0$, then $\rho_l = l - 1$ and N_l is the set of all pairs $(i, l+2-i)$, $i = 1, 2, \ldots, l+1$. So $d(l) = \nu_l = l+1$ for all $l \in \mathbb{N}$.

Theorem 5.24 *Let* $D(l) = \{ \ (x,y) \mid \ x \text{ and } y \text{ are gaps and } x + y = \rho_{l+1} \ \}$. *Then*

$$\nu_l = l + 1 - g(l+1) + \#D(l),$$

where $g(l+1) = g$ *if* $l \geq c - g$ *and* $\#D(l) = 0$ *if* $l > 2c - g - 2$. *Furthermore* $d(l) \geq d_G(l) = l + 1 - g$ *and equality holds if* $l > 2c - g - 2$.

Proof.

(1) For a given integer l, define the following sets. $A(l)$ is the set of pairs of nonnegative integers (x,y) such that $x + y = \rho_{l+1}$. Let $B(l)$ be the set of pairs $(x,y) \in A(l)$ such that x is a gap, and let $C(l)$ be the set of pairs

$(x, y) \in A(l)$ such that y is a gap. Clearly $A(l) = N_l \cup B(l) \cup C(l)$ and $D(l) = B(l) \cap C(l)$ and N_l is disjoint from $B(l) \cup C(l)$. Hence

$$\nu_l = \#A(l) - \#B(l) - \#C(l) + \#D(l).$$

The number of elements of $A(l)$ is $\rho_{l+1} + 1$. Let $x \in \mathbb{N}_0$. Then x is a gap smaller than ρ_{l+1} if and only if there exists a unique y such that $(x, y) \in B(l)$. So $\#B(l) = g(l + 1)$, and similarly $\#C(l) = g(l + 1)$. The equality for ν_l follows now, since $g(l + 1) = \rho_{l+1} - l$ by Lemma 5.6.

(2) If $l \geq c - g$, then $g(l + 1) = g$ by Lemma 5.6 (3).

(3) Now suppose that $l > 2c - g - 2$. If $g = 0$, then $\nu_l = l + 1$ for all $l \in \mathbb{N}$ by Example 5.23. So we may assume that $g > 0$ and $c \geq 2$. So $2c - g - 2 \geq c - g$. Hence $\rho_{l+1} = l + g > 2c - 2$. Let x and y be gaps. Then $x, y \leq c - 1$. If moreover $x + y = \rho_{l+1}$, then $\rho_{l+1} \leq 2c - 2$. Therefore such a pair (x, y) does not exist. So $D(l)$ is empty if $l > 2c - g - 2$.

(4) The statement about $d(l)$ and $d_G(l)$ follows immediately from the definitions and the above results on ν_l. □

Example 5.25 Let us illustrate the theorem by means of a diagram of the semigroup of the Klein quartic. Put the elements of the semigroup in an array with $x + y$ on the entry (x, y) if x and y are nongaps, and a dot otherwise. Look at the diagonal $x + y = \rho_{l+1}$ and count the the number of times we see ρ_{l+1}. In this way we determine ν_l, the size of N_l. The dots in the rows correspond to elements of the set $B(l)$, and the dots in the columns to elements in $C(l)$. An element (x, y) of $D(l)$ corresponds with the x-th column with dots intersecting the y-th row with dots. If l is large enough, then they do not intersect anymore.

0	·	·	3	·	5	6	7	8	9	10	11	12
·	·	·	·	·	·	·	·	·	·	·	·	
·	·	·	·	·	·	·	·	·	·	·		
3	·	·	6	·	8	9	10	11	12			
·	·	·	·	·	·	·	·	·				
5	·	·	8	·	10	11	12					
6	·	·	9	·	11	12						
7	·	·	10	·	12							
8	·	·	11	·								
9	·	·	12									
10	·	·										
11	·											
12												

This gives again the values of the table of Example 4.16.

l	1	2	3	4	5	6	7	8	9
ρ_{l+1}	3	5	6	7	8	9	10	11	12
ν_l	2	2	3	2	4	4	5	6	7

We will compare the redundancy of the codes C_l with the codes $\tilde{C}(d)$ and $\tilde{C}_\varphi(d)$ of 4.22.

Proposition 5.26 *Let ρ be a weight function. Let g be the number of gaps of the corresponding semigroup. Then $r_\varphi(d) \leq r(d) \leq d - 1 + g$.*

Proof. The inequality $r_\varphi(d) \leq r(d)$ follows from the inclusion $R_\varphi(d) \subseteq R(d)$.

If $l \in \mathbb{N}_0$ and $l \geq d - 1 + g$, then $\nu_l \geq l + 1 - g \geq d$, by Theorem 5.24. So $l + 1 \notin R(d)$. Hence $R(d) \subseteq \{1, 2, \ldots, d - 1 + g\}$ and $r(d) \leq d - 1 + g$. □

The following technical lemma is needed in order to say more in the two generator case.

Lemma 5.27 *Suppose a and b are two positive integers which are relatively prime and such that $a > b$. Let Λ be the semigroup generated by a and b. Let $\rho_{l+1} = xb + ya$ for some nonnegative integers x and y. If $\rho_{l+1} < (b-1)a$, then $\nu_l = (x+1)(y+1)$ and there is at least one gap in the interval $[\rho_{l+1} - \nu_l, \rho_{l+1}]$.*

Proof. Let $m = \rho_{l+1}$. Let $\nu = \nu_l$. Then ν is the number of pairs $(m_1, m_2) \in \Lambda^2$ such that $m_1 + m_2 = m$.

We will use several times the fact that if $m' \in \Lambda$ and $m' < (b-1)a$, then $y < b$, so there exist uniquely determined nonnegative integers x and y such that $m = xb + ya$, since $\gcd(a, b) = 1$.

(1) Let $(i, j) \in \mathbb{N}_0^2$ such that $0 \le i \le x$ and $0 \le j \le y$. Define $m_1(i,j) = ib + ja$ and $m_2(i,j) = (x-i)b + (y-j)a$. Then $m_1(i,j)$, $m_2(i,j) \in \Lambda$ and $m_1(i,j) + m_2(i,j) = m$. The $m_1(i,j)$ are mutually distinct. Hence $\nu \ge (x+1)(y+1)$.

If (m_1, m_2) is a pair such that $m = m_1 + m_2$ and $m_1, m_2 \in \Lambda$, then $m_1 = x_1 b + y_1 a$ and $m_2 = x_2 b + y_2 a$ for some nonnegative integers x_1, y_1, x_2 and y_2. So $xb + ya = (x_1 + x_2)b + (y_1 + y_2)a$. Hence $x_1 + x_2 = x$ and $y_1 + y_2 = y$. Therefore $m_t = m_t(x_t, y_t)$ for $t = 1, 2$. So $\nu = (x+1)(y+1)$.

(2) Let $m - i$ be an element of the interval $[m - \nu, m]$. Write

$$m - i = x_i b + y_i a, \quad 0 \le y_i < b \quad \text{for } i = 0, 1, \ldots, \nu.$$

If we can show that one of the x_i's is negative, then there is at least one gap in $[m - \nu, m]$. Consider two cases:

(2.i) $\nu < b$. Here the y_i, $i = 0, \ldots, \nu$ are $\nu+1$ distinct nonnegative integers. So there is at least one $y_i \ge \nu = (x+1)(y+1)$. For the corresponding x_i we have

$$x_i b = m - i - y_i a \le xb + ya - i - (x+1)(y+1)a \le x(b-a) < 0,$$

since $b < a$.

(2.ii) $\nu \ge b$. Then $m - i$ takes on all possible values modulo b. Furthermore $m - i \equiv a y_i \pmod b$ and $\gcd(a, b) = 1$. Hence y_i takes on all possible values modulo b. So we find $y_i = b - 1$ for some $i = 0, 1, \ldots, \nu$. For the corresponding x_i we have

$$x_i b = m - i - y_i a \le m - (b-1)a < 0,$$

since it is assumed that $m < (b-1)a$.

In both cases it is shown that one of the x_i's is negative. □

Proposition 5.28 *Let the semigroup of the weight function be generated by a and b such that $b < a$ and $\gcd(a, b) = 1$. Let (ρ_i) be an enumeration of the semigroup in increasing order. Then*

$$d(l) = j + 1 \qquad \textit{if } l < g \textit{ and } (j-1)a < \rho_{l+1} \le ja.$$

Proof. The semigroup is symmetric, so $c = 2g$, and $c = (a-1)(b-1)$ by Proposition 5.11. Let $l < g$. Then $l < c - g$, so $\rho_{l+1} < c - 1$ by Lemma 5.6. Hence $\rho_l < (b-1)a$. Write $\rho_{l+1} = xb + ya$ for some nonnegative integers x and y. Then $\nu_l = (x+1)(y+1)$ by Lemma 5.27. If moreover $\rho_{l+1} = ja$, then $x = 0$ and $y = j$, so $\nu_l = j + 1$. Now assume $(j-1)a < \rho_{l+1} = xb + ya < ja$. So $0 \leq y \leq j - 1$. Then $\nu_l = (x+1)(y+1)$ is strictly larger than

$$\left(\frac{(j-1-y)a}{b} + 1\right)(y+1) \geq (j-1-y) + (y+1) = j.$$

Hence $d(l) = \min\{\ \nu_m \mid m \geq l\ \} = j + 1$. □

Example 5.29 This is a continuation of Example 4.17 with the Hermitian curve over $\mathbb{F}_{16}$. We have seen that the seqence ν_l is given by:

l	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
ν_l	2	2	3	4	3	4	6	6	4	5	8	9	8	9	10	12	12

and $\nu_l = l - 5$ for all $l \geq 17$. Notice that the ν_l are nongaps for all $l \geq 6$. Furthermore if $6 \leq l \leq 9$, then $\nu_l \geq 4$ and 4 is the smallest nongap which is at least $l - 5$.

Theorem 5.30 *Let the semigroup of the weight function be generated by two elements. If $l \geq g$, then*

$$d(l) = min\{\ \rho_t \ \mid \ \rho_t \geq l + 1 - g\ \}.$$

Proof. The semigroup Λ is symmetric by Proposition 5.11. So $c = 2g$. Let $l \geq g = c - g$. Then $\rho_{l+1} = l + g$ by Lemma 5.6.

(1) If $l > 3g - 2 = 2c - g - 2$, then $d(l) = l + 1 - g$ by Theorem 5.24. Furthermore $l - 2g + 2 > g = c - g$. So $\rho_{l-2g+2} = l + 1 - g$. Hence $d(l) = \rho_{l-2g+2}$.

(2) Suppose $g \leq l \leq 3g - 2$. Then $\rho_{l+1} = l + g$. So $2g \leq \rho_{l+1} \leq 4g - 2$. We may write $\rho_{l+1} = 2g - 1 + \kappa$, $1 \leq \kappa \leq 2g - 1$. So $\kappa = l + 1 - g$. The number ν_l is equal to $l + 1 - g + \#D(l)$, by Theorem 5.24. For the estimation of the number of elements of $D(l)$ we will consider two cases:

(2.*i*) κ is a nongap. Let $(x, y) \in D(l)$. Then x and y are gaps and $\rho_{l+1} = x + y$. So $(2g - 1 - x) + \kappa = y$. The semigroup is symmetric, so $2g - 1 - x$ is a nongap. The sum of two nongaps is of course again a nongap.

Hence y cannot be a gap and $D(l)$ is empty. Therefore $\nu_l = l+1-g = \kappa$ is a nongap ρ_t for some t.

(2.*ii*) κ is a gap. Now there is a t such that $\rho_{t-1} < \kappa < \rho_t$. There exists an $L \geq l$ such that $\rho_{L+1} = 2g-1+\rho_t$. By the argument in (2.*i*) we have that $\nu_L = L+1-g = \rho_t$. We will show that $\nu_l \geq \rho_t$.

The function $\#D(l)$ is defined by a condition on gaps. But for symmetric semigroups such a condition can be translated into a condition on nongaps. Define $x' = 2g-1-x$ if $x \in \mathbb{N}_0$, $0 \leq x \leq 2g-1$. Then

$$x, y \in (\mathbb{N}_0 \setminus \Lambda),\ x+y = \rho_{l+1} \text{ if and only if } x', y' \in \Lambda,\ x'+y' = 2g-1-\kappa.$$

So $2g-1-\kappa$ is a nongap ρ_{u+1} and the number of elements of $D(l)$ is equal to ν_u. Hence there is a gap in the interval $[\rho_{u+1}-\nu_u, \rho_{u+1}]$ by Lemma 5.27. Recall that $\rho_{t-1} < \kappa < \rho_t$. So the numbers $\kappa, \kappa+1, \ldots, \rho_t-2, \rho_t-1$ are all gaps. So

$$2g-\rho_t, 2g+1-\rho_t, \ldots, 2g-2-\kappa, 2g-1-\kappa = \rho_{u+1}$$

are all nongaps. So $\rho_{u+1} - \nu_u < 2g-\rho_t$. But $\rho_{u+1} = 2g-1-\kappa$ by definition. Therefore $\rho_t - \kappa \leq \nu_u = \#D(l)$. Hence $\nu_l = l+1-g+\#D(l) \geq \rho_t$, since $\kappa = l+1-g$. This implies $d(l) = \rho_t$ is the smallest nongap which is at least $l+1-g$. □

5.3 Hermitian codes

In this section the theory will be applied to the Hermitian codes. This is a continuation of Example 5.29 on the Hermitian curve. The length of these codes is $n = r^3$, and the number of gaps is $g = (r^2-r)/2$.

The set $\{\ x^\alpha y^\beta \mid 0 \leq \beta < r\ \}$ is used as a basis for the ring R and $\rho(x^\alpha y^\beta) = \alpha r + \beta(r+1)$. If $\alpha \geq q$, then $ev_{\mathcal{P}}(x^\alpha y^\beta) = ev_{\mathcal{P}}(x^{\alpha-q} y^\beta)$. So the set

$$\{\ ev_{\mathcal{P}}(x^\alpha y^\beta) \mid 0 \leq \alpha < q,\ 0 \leq \beta < r\ \}$$

generates $\mathbb{F}_q^n$ and has $qr = n$ elements. So it is a basis.

The nongaps are of the form $\rho_l = \alpha r + \beta(r+1) = (\alpha+\beta)r + \beta$, where α and β are nonnegative integers such that $\beta < r$. If $1 \leq \alpha < r$, then the integers of the interval $[(\alpha-1)(r+1)+1, \alpha r - 1]$ are gaps. Furthermore every gap lies in such an interval.

In order to use the improved bound $d_{\mathcal{P}}(l)$ those l with $C_l = C_{l+1}$ must be determined. If $l \leq n - g$, then $\rho_l < n$, so E_l has dimension l, by Theorem 5.18, and C_l has dimension $n - l$. So $C_l \neq C_{l+1}$ and $d(l) = d_{\mathcal{P}}(l)$ for all $l < n - g$. If $l \geq n - g$, then $l > 3g - 2$, so ν_l is strictly increasing. Therefore, although $d_{\mathcal{P}}(l) > d(l)$ for some values of l, there always exists an $m > l$ such that $C_l = C_m$ and $d_{\mathcal{P}}(l) = d(m)$.

Five cases will be distinguised to determine the minimum distance of C_l.

(1) If $l > 3g - 2$, then Theorem 5.24 gives $d(l) = l + 1 - g$.

(2) If $g \leq l \leq 3g - 2$, then Theorem 5.30 can be used. Write $l = 3g - 1 - (x - 1)r - y$ with $0 \leq y < r$ and $1 \leq x < r$. The smallest nongap ρ_t has to be determined such that $\rho_t \geq l + 1 - g = (r - x - 1)r + (r - y)$. If $x < y$, then $(r - x - 1)r + (r - y) = (i + j)r + j$ with $i = y - x - 1$ and $j = r - y$. So the minimum is $(r - x)r - y$ and therefore $d(l) = (r - x)r - y$. If $x \geq y$, then $(r - x - 1)r + (r - y)$ is in the interval $[(i - 1)(r + 1) + 1, ir - 1]$ with $i = r - x$. So $d(l) = (r - x)r$ in this case.

(3) Suppose $l < g$. Write $l = u(u + 1)/2 + (v + 1)$ with $0 \leq v \leq u < r - 1$. Then $\rho_l = ur + v$. This gives $\rho_{l+1} = ur + v + 1$ if $v < u$ and $\rho_{l+1} = (u + 1)r$ if $v = u$. Therefore $(u - 1)(r + 1) < \rho_{l+1} < u(r + 1)$ in the first case and $u(r + 1) < \rho_{l+1} < (u + 1)(r + 1)$ in the second case. Proposition 5.28 implies that $d(l)$ is equal to $u + 1$ and $u + 2$, respectively.

(4) Suppose $l \geq n - g$. Write $l = n - g + \alpha r + \beta$ with $0 \leq \beta < r$. Then $\rho_{l+1} = l + g = n + \alpha r + \beta = r(q + (\alpha - \beta) + \beta) + \beta$ which means that $f_{l+1} = x^{q+(\alpha-\beta)}y^{\beta}$. If $\beta \leq \alpha$, then the exponent of x is at least q, so $ev_{\mathcal{P}}(f_{l+1}) \in E_l$ and $C_l = C_{l+1}$. If $\alpha < \beta$, then $ev_{\mathcal{P}}(f_{l+1})$ is an element of the chosen basis, so $C_l \neq C_{l+1}$. Therefore

$$d_{\mathcal{P}}(l) = \min\{\ n + \gamma r + \delta + 1 - g \mid 0 \leq \gamma < \delta < r,\ l \leq n + \gamma r + \delta\ \}.$$

This minimum is $n + \alpha r + \beta + 1 - 2g$ if $\alpha < \beta$, and $n + \alpha r + (\alpha + 1) + 1 - 2g$ if $\beta \leq \alpha$.

(5) If $l > n + g$, then $C_l = 0$, since $x^{q-1}y^{r-1}$ is the function which evaluates to the last element of the basis for $\mathbb{F}_q^n$, and $(q - 1)r + (r - 1)(r + 1) = (n + g) - 1 + g$. So this function is the $(n + g)$-th element of the basis of R.

For the Hermitian codes, the above lower bounds give in fact the true minimum distance. We refer to the literature in the Notes for this fact. One needs a result similar to the RM codes, stating that the codes C_l are also of the form $E_{l^\perp}$.

The values of the Hermitian codes are summarized in the following table.

l	$d(C_l)$	
$1 \leq l < g$		$l = \binom{u+1}{2} + v + 1$, $v \leq u < r - 1$
	$u + 1$	if $v < u$
	$u + 2$	if $v = u$
$g \leq l \leq 3g - 2$		$l = 3g - 1 - (x - 1)r - y$, $y < r$
	$(r - x)r - y$	if $x < y$
	$(r - x)r$	if $x \geq y$
$3g - 2 < l < n - g$	$l + 1 - g$	
$n - g \leq l \leq n + g$		$l = n - g + \alpha r + \beta$, $\beta < r$
	$n + \alpha r + \beta + 1 - 2g$	if $\alpha < \beta$
	$n + \alpha r + \alpha + 2 - 2g$	if $\alpha \geq \beta$

5.4 Telescopic semigroups

Definition 5.31 Let $(a_1, \ldots, a_k)$ be a sequence of positive integers with greatest common divisor 1. Define

$$d_i = \gcd(a_1, \ldots, a_i) \text{ and } A_i = \{a_1/d_i, \ldots, a_i/d_i\}$$

for $i = 1, \ldots, k$. Let $d_0 = 0$. Let Λ_i be the semigroup generated by A_i. If $a_i/d_i \in \Lambda_{i-1}$ for $i = 2, \ldots, k$, then the sequence $(a_1, \ldots, a_k)$ is called *telescopic*. A semigroup is called telescopic if it is generated by a telescopic sequence.

Remark 5.32 If $(a_1, \ldots, a_k)$ is telescopic, then $\gcd(a_1/d_i, \ldots, a_i/d_i) = 1$ and the sequence $(a_1/d_i, \ldots, a_i/d_i)$ is telescopic for $i = 2, \ldots, k$.

If $d_i = 1$ for a telescopic sequence $(a_1, \ldots, a_k)$, then $(a_1, \ldots, a_i)$ is also telescopic and generates the same semigroup.

Example 5.33 Semigroups generated by two relatively prime elements are telescopic. The sequence $(4, 6, 5)$ is telescopic, since $d_2 = 2$ and 5 is an element of the semigroup generated by $4/2$ and $6/2$. The sequence $(4, 5, 6)$ is not telescopic.

Lemma 5.34 *If $(a_1, \ldots, a_k)$ is telescopic and $m \in \Lambda_k$, then there exist uniquely determined nonnegative integers $x_1, x_2, \ldots, x_k$ such that $0 \le x_i < d_{i-1}/d_i$ for $i = 2, \ldots, k$ and*

$$m = \sum_{i=1}^{k} x_i a_i.$$

This representation is called the normal representation *of m by $(a_1, \ldots, a_k)$.*

Proof. The proof is by induction on the number k of entries in the sequence. For $k = 1$ there is nothing to prove. For $k = 2$ the Lemma says: if $\gcd(a_1, a_2) = 1$, then every $m \in \Lambda_2$ can be written uniquely as $m = x_2 a_2 + x_1 a_1$, $0 \le x_2 < a_1$. In fact this property was already used in the proof of Proposition 5.11. Now suppose the lemma is proven for all telescopic sequences with $k-1$ entries and look at $m \in \Lambda_k$. There exist $x_k \in \mathbb{N}_0$ and $u \in \Lambda_{k-1}$ such that $m = x_k a_k + d_{k-1} u$, since $\Lambda_k = \langle a_k \rangle + d_{k-1}\Lambda_{k-1}$. Write $x_k = w d_{k-1} + v$, $0 \le v < d_{k-1}$ and get $m = v a_k + d_{k-1}(u + w a_k)$. Now $a_k \in \Lambda_{k-1}$, and therefore $u + w a_k \in \Lambda_{k-1}$. Remember that also $(a_1/d_{k-1}, \ldots, a_{k-1}/d_{k-1})$ is telescopic by Remark 5.32. Let $d'_i = \gcd(a_1/d_{k-1}, \ldots, a_i/d_{k-1})$ for $i = 1, \ldots, k-1$. Then there exist $0 \le x_i < d'_{i-1}/d'_i$ for $i = 2, \ldots, k-1$, such that

$$u + w a_k = \sum_{i=1}^{k-1} x_i \frac{a_i}{d_{k-1}}$$

is a normal representation by $(a_1/d_{k-1}, \ldots, a_{k-1}/d_{k-1})$. Therefore $m = v a_k + \sum_{i=1}^{k-1} x_i a_i$ is a normal representation by $(a_1, \ldots, a_k)$, since $d'_{i-1}/d'_i = d_{i-1}/d_i$.

For the uniqueness assume m has two normal representations $\sum_{i=1}^{k} x_i a_i = m = \sum_{i=1}^{k} y_i a_i$, where $0 \le x_i, y_i < d_{i-1}/d_i$ for $i = 2, \ldots, k$. Let l be the largest index for which $x_i \ne y_i$. Then $\sum_{i=1}^{l} x_i a_i = \sum_{i=1}^{l} y_i a_i$ and $(x_l - y_l) a_l = \sum_{i=1}^{l-1} (y_i - x_i) a_i$. Hence the right-hand side is a multiple of d_{l-1} and $\gcd(a_l/d_l, d_{l-1}/d_l) = 1$, so $x_l - y_l$ is a nonzero multiple of d_{l-1}/d_l which gives a contradiction. □

Proposition 5.35 *Let Λ_k be the semigroup generated by the telescopic sequence $(a_1, \ldots, a_k)$. Then*

$$c(\Lambda_k) - 1 = d_{k-1}(c(\Lambda_{k-1}) - 1) + (d_{k-1} - 1) a_k = \sum_{i=1}^{k} (d_{i-1}/d_i - 1) a_i,$$

$$g(\Lambda_k) = d_{k-1}g(\Lambda_{k-1}) + (d_{k-1} - 1)(a_k - 1)/2 = c(\Lambda_k)/2.$$

So telescopic semigroups are symmetric. Here we put $d_0 = 0$.

Proof. If $k = 1$, then $\Lambda_1 = \mathbb{N}_0$. So the conductor is 0 and the number of gaps is 0. This is in accordance with the formulas. For $k = 2$ we get Proposition 5.11.

Assume $k > 1$. Since $\gcd(a_k, d_{k-1}) = 1$ every integer $m \in \mathbb{N}_0$ can be uniquely represented as $m = va_k + d_{k-1}w$, $0 \leq v < d_{k-1}$. Here w may be negative. So by Lemma 5.34 the gaps of Λ_k are exactly the numbers m, where the corresponding w is either a gap of Λ_{k-1} or w is negative. Hence the first equation, involving the largest gap $c(\Lambda_k)-1$ in terms of the conductor $c(\Lambda_k)$, follows immediately, and for the second, we proceed by induction.

For every value of $0 \leq v < d_{k-1}$ there are $g(\Lambda_{k-1})$ gaps of Λ_k coming from those of Λ_{k-1}. In addition we get the gaps of the form $m = va_k + d_{k-1}w$, where $w < 0$. But these are exactly the gaps of the semigroup $< a_k, d_{k-1} >$, the number of which we know to be $(d_{k-1}-1)(a_k-1)/2$, by Proposition 5.11. So the total number of gaps is equal to $d_{k-1}g(\Lambda_{k-1}) + (d_{k-1} - 1)(a_k - 1)/2$. The remaining result on the symmetry now follows by induction. □

Example 5.36 For the ideal considered in Example 3.22 the semigroup of nongaps is generated by the numbers $a_1a_2 \cdots a_{i-1}b_i \cdots b_{m-1}$ with $1 \leq i \leq m$, where the empty product is equal to 1 by definition. The greatest common divisor of the first i elements of these numbers is equal to $d_i = b_i \cdots b_{m-1}$, since $gcd(a_i, b_j) = 1$ for all $i \leq j$. So Λ_i is generated by the numbers

$$a_1a_2 \cdots a_{j-1}b_j \cdots b_{i-1}, \quad 1 \leq j \leq i,$$

and the semigroup is telescopic. Proposition 5.35 implies that

$$2g - 1 = c - 1 = \sum_{i=2}^{m}(b_{i-1} - 1)a_1 \cdots a_{i-1}b_i \cdots b_{m-1} - b_1b_2 \cdots b_{m-1},$$

where $g = g(\Lambda_m)$ and $c = c(\Lambda_m)$.

5.5 Notes

The connection between properties of semigroups and the minimum distance of algebraic geometry codes was made in [32, 33] and [53].

The history of the notion of telescopic semigroups and generalizations of Theorems 5.30 and 5.28 for these semigroups can be found in [53]. See also [4].

Hermitian codes have been studied extensively by many authors; their true minimum distance was determined in [57].

The formulation of Theorem 5.30 is from [49], which is an extension of results obtained in [69].

The formula for the genus in Example 5.36 can be found in [25] and a special case in [77].

6 Decoding algebraic geometry codes

In this section we treat two decoding algorithms for the codes C_l presented in Sections 4 and 5. The basic algorithm corrects up to $\lfloor (d_G(l) - 1 - g)/2 \rfloor$ errors when ρ is a weight function with g gaps. An extended algorithm using majority voting on unknown syndromes enables one to decode up to half the order bound if ρ is an arbitrary order function.

6.1 The decoding problem

Let C be a linear code in $\mathbb{F}_q^n$ of minimum distance d. If $\mathbf{c}$ is a transmitted word and $\mathbf{c} + \mathbf{e}$ is the received word, then we call $\mathbf{e}$ the *error vector* and $\{i | e_i \neq 0\}$ the set of *error positions*. The e_i's are called the *error values* and $wt(\mathbf{e})$ is the *number of errors* of the received word. If $\mathbf{y}$ is the received word and the distance of $\mathbf{y}$ to the code C is t', then there exists a codeword $\mathbf{c}'$ and an error vector $\mathbf{e}'$ such that $\mathbf{y} = \mathbf{c}' + \mathbf{e}'$ and $wt(\mathbf{e}') = t'$. If the number of errors is at most $(d-1)/2$, then we are sure that $\mathbf{c} = \mathbf{c}'$ and $\mathbf{e} = \mathbf{e}'$. In other words, the nearest codeword to $\mathbf{y}$ is unique when $\mathbf{y}$ has distance at most $(d-1)/2$ to C.

Define $C^* = C \cup \{?\}$. A map

$$\mathcal{D} : \mathbb{F}_q^n \longrightarrow C^*$$

is called a *decoder* for the code C if $\mathcal{D}(\mathbf{c}) = \mathbf{c}$, for all $\mathbf{c} \in C$. We allow the decoder to give as outcome "?", when it fails to find a codeword. A *maximum likelihood decoder* for a code C is a decoder $\mathcal{D}$ such that $\mathcal{D}(\mathbf{y})$ is a closest codeword to $\mathbf{y}$ for all $\mathbf{y}$. A decoder $\mathcal{D}$ for a code C is called a *bounded distance* decoder that *corrects t errors* if $\mathcal{D}(\mathbf{y})$ is a nearest codeword for all $\mathbf{y} \in \mathbb{F}_q^n$ such that $d(\mathbf{y}, C) \leq t$. A decoder $\mathcal{D}$ for a code C of minimum distance d *decodes up to half the minimum distance* if $\mathcal{D}(\mathbf{y})$ is the nearest codeword for all $\mathbf{y} \in \mathbb{F}_q^n$ such that $d(\mathbf{y}, C) \leq (d-1)/2$.

Concerning statements about the the number of additions and multiplications we use the "big $\mathcal{O}$" notation. We say $f(n) = \mathcal{O}(g(n))$ for $n \to \infty$ if and only if there exists a positive constant c and an integer n_0 such that $|f(n)| \leq c|g(n)|$ for all $n \geq n_0$. Of course, for many classes of codes, "$n \to \infty$" makes no sense. An algorithm has *polynomial complexity* if the number of operations is a polynomial in the length of the input n. Concerning decoding algorithms, the received word is the input, so the length of the code is a measure of the input. The known decoding algorithms which have polynomial complexity decode only up to a certain bound, for instance up to half the (designed) minimum distance. All decoding algorithms for algebraic geometry codes, that we will treat, decode up to half some designed minimum distance and have complexity $\mathcal{O}(n^3)$ or less for $n \to \infty$.

Errors can be corrected by solving a system of linear equations involving syndromes, see 4.6, if we have complete information about the error positions, in other words if we have *erasures* only.

Proposition 6.1 *Let C be a linear code in $\mathbb{F}_q^n$ with parity check matrix $\mathbf{H}$. Suppose we have a received word $\mathbf{y}$ with error vector $\mathbf{e}$ and know a set J with at most $d(C) - 1$ elements that contains the set of error positions. Then the error vector $\mathbf{e}$ is the unique solution of the following linear equations:*

$$\mathbf{x}\mathbf{H}^T = \mathbf{y}\mathbf{H}^T \quad \textit{and} \quad x_j = 0 \ \textit{ for all } \ j \notin J.$$

Proof. It is clear that the error vector is a solution. If $\mathbf{x}$ is another solution, then $(\mathbf{x} - \mathbf{e})\mathbf{H}^T = 0$. Therefore $\mathbf{x} - \mathbf{e}$ is an element of C, and moreover it is supported at J. So its weight is at most $d(C) - 1$. So it must be zero. Therefore $\mathbf{x} = \mathbf{e}$. □

Thus we have shown that we can reduce error decoding to the problem of finding the error positions. If we want to decode all received words with t errors, then there are $\binom{n}{t}$ possible t-sets of error positions one has to consider. This number grows exponentially in n when t/n tends to a nonzero real number. From Proposition 6.1 it is enough to find an $(n, d-1, t)$ *covering system.* That is to say a collection $\mathcal{J}$ of subsets J of $\{1, \ldots, n\}$, such that all $J \in \mathcal{J}$ have $d-1$ elements and every subset of $\{1,\dots,n\}$ of size t is contained in at least one $J \in \mathcal{J}$. The size of such a covering system is considerably smaller than the number of all possible t-sets, but is at least $\binom{n}{t}/\binom{d-1}{t}$. This number also grows exponentially in n.

6.2 The basic algorithm

Suppose that we are in the situation of Section 4 with an order function ρ on an affine $\mathbb{F}_q$-algebra R and a surjective algebra morphism $\varphi : R \to \mathbb{F}_q^n$. Let $\{f_i \mid i \in \mathbb{N}\}$ be a basis of R over $\mathbb{F}_q$ such that $\rho(f_i) < \rho(f_{i+1})$ for all $i \in \mathbb{N}$. Let L_l be the vector space with $f_1, \dots, f_l$ as a basis. We defined $l(i,j)$ as the smallest positive integer l such that $f_i f_j \in L_l$. Let $\mathbf{h}_i = \varphi(f_i)$. Let $C_l = \{\mathbf{c} \in \mathbb{F}_q^n \mid \mathbf{c}\cdot\mathbf{h}_i = 0 \ \text{ for all } \ i \leq l\}$ as before. So $\mathbf{h}_i * \mathbf{h}_j$ is a parity check for C_l if $l(i,j) \leq l$. The syndromes were defined 4.6 by $s_i(\mathbf{y}) = \mathbf{y}\cdot\mathbf{h}_i$ and $s_{ij}(\mathbf{y}) = \mathbf{y}\cdot(\mathbf{h}_i * \mathbf{h}_j)$. Define the $i \times j$ submatrix $\mathbf{S}(i,j)$ of the matrix of syndromes $\mathbf{S}(\mathbf{y})$, as defined in 4.6, by

$$\mathbf{S}(i,j) = (s_{i',j'}(\mathbf{y}) \mid 1 \leq i' \leq i, 1 \leq j' \leq j).$$

Suppose that we want to correct the errors of words with respect to the code C_l. If $\mathbf{y}$ is a received word and $\mathbf{y} = \mathbf{c} + \mathbf{e}$ with $\mathbf{c} \in C_l$, then $s_{i,j}(\mathbf{y}) = s_{i,j}(\mathbf{e})$ for all i,j such that $l(i,j) \leq l$.

Definition 6.2 Assume that $l(i,j) \leq l$. Let $\mathbf{y} \in \mathbb{F}_q^n$. Define the space

$$K_{ij}(\mathbf{y}) = \{\ f \in L_j \mid \mathbf{y}\cdot\varphi(fg) = 0 \text{ for all } g \in L_i\}.$$

Then $K_{ij}(\mathbf{y})$ is a subspace of L_j and it is the kernel of the linear map $L_j \to L_i$ with matrix $\mathbf{S}(i,j)$ with respect to the bases $f_1, \dots, f_j$ and $f_1, \dots, f_i$ of L_j and L_i, respectively. Hence $K_{ij}(\mathbf{y}) = K_{ij}(\mathbf{e})$.

Definition 6.3 Let J be a subset of $\{1, \dots n\}$. Define the subspace

$$L_j(J) = \{\ f \in L_j \mid \varphi(f)_k = 0 \ \text{ for all } \ k \in J\ \},$$

where $\varphi(f)_k$ denotes the k-th coordinate of $\varphi(f)$.

Lemma 6.4 *If* $I = \mathrm{supp}(\mathbf{e}) = \{k \in \{1, \ldots, n\} \mid e_k \neq 0\}$*, then* $L_j(I) \subseteq K_{ij}(\mathbf{y})$*. If moreover* $d(C_i) > \mathrm{wt}(\mathbf{e})$*, then* $L_j(I) = K_{ij}(\mathbf{y})$*.*

Proof. Let $f \in L_j(I)$. Then $\varphi(f)_k = 0$ for all k such that $e_k \neq 0$, and therefore

$$\mathbf{e} \cdot (\varphi(f) * \varphi(g)) = \sum_{e_k \neq 0} e_k(\varphi(f) * \varphi(g))_k = 0$$

for all $g \in L_i$. So $f \in K_{ij}(\mathbf{e}) = K_{ij}(\mathbf{y})$.

Suppose moreover that $d(C_i) > \mathrm{wt}(\mathbf{e})$. Let $f \in K_{ij}(\mathbf{y})$ and let $\mathbf{a} = \varphi(f)$, then $f \in K_{ij}(\mathbf{e})$ and hence

$$(\mathbf{e} * \mathbf{a}) \cdot \varphi(g) = \mathbf{e} \cdot (\varphi(f) * \varphi(g)) = 0$$

for all $g \in L_i$, giving $\mathbf{e}*\mathbf{a} \in C_i$. Now $\mathrm{wt}(\mathbf{e}*\mathbf{a}) \leq \mathrm{wt}(\mathbf{e}) < d(C_i)$ and therefore $\mathbf{e}*\mathbf{a} = 0$ meaning that $e_k\varphi(f)_k = 0$ for all $k \in \{1, 2, \ldots, n\}$. Hence $\varphi(f)_k = 0$ for all $k \in I = \mathrm{supp}(\mathbf{e})$ and therefore $f \in L_j(I)$. □

Let I be the set of error positions supp($\mathbf{e}$). The set of zero coordinates of $\varphi(f)$, where $f \in L_j(I)$ contains the set of error positions. For that reason the elements of $L_j(I)$ are called *error-locator* functions. But the space $L_j(I)$ is not known. The space $K_{ij}(\mathbf{y})$ can be computed after receiving the word $\mathbf{y}$. The equality $L_j(I) = K_{ij}(\mathbf{y})$ implies that all elements of $K_{ij}(\mathbf{y})$ are error-locator functions.

More generally, every element f of R satisfying $\varphi(f)_k = 0$ for all $k \in \mathrm{supp}(\mathbf{e})$ is called an *error-locator* and the error-locators obviously constitute an ideal L of R. If $\mathrm{wt}(\mathbf{e}) = t$ then the dimension of R/L as an $\mathbb{F}_q$-vector space is t.

Suppose $l(i,j) \leq l$. The *basic algorithm* $\mathcal{A}(i,j)$ for the code $C = C_l$ computes the kernel $K_{ij}(\mathbf{y})$ for every received word $\mathbf{y}$. If this kernel is nonzero, it takes a nonzero element f and determines the set J of zero positions of f. If $d(C_i) > wt(\mathbf{e})$, where $\mathbf{e}$ is the error-vector, then J contains the support of $\mathbf{e}$ by Lemma 6.4. If the set J is not too large, Proposition 6.1 can be applied to get the error values.

Thus we have a basic algorithm for every pair (i,j) such that $l(i,j) \leq l$. If j is too small with respect to the number of errors, then $K_{i,j}(\mathbf{y}) = 0$. If j is large, then i becomes small, which results in a large code C_i, and it will be difficult to meet the requirement $d(C_i) > wt(\mathbf{e})$.

Proposition 6.5 *Let ρ be a weight function with g gaps. Then the basic algorithm corrects $\lfloor (d_G(l) - 1 - g)/2 \rfloor$ errors for the code C_l with complexity $\mathcal{O}(n^3)$.*

Proof. We may assume $t = \lfloor (d_G(l) - 1 - g)/2 \rfloor \geq 1$, so $l \geq 2g + 2$ and $\rho_l = l + g - 1$. Assume for simplicity that l is even. The Goppa designed minimum distance $d_G(l) = l + 1 - g$. So $t = l/2 - g$. Let $j = t + 1$ and let $i = l/2$. Then $\rho_j \leq l/2$ and $\rho_i = l/2 + g - 1$. So $\rho_i + \rho_j \leq l + g - 1 \leq \rho_l$. So $l(i,j) \leq l$ and the basic algorithm $\mathcal{A}(i,j)$ can be applied to decode C_l.

If a received word $\mathbf{y}$ has at most t errors, then the error vector $\mathbf{e}$ with support I has size at most t and $L_j(I)$ is not zero, since I imposes at most t linear conditions on L_j and the dimension of L_j is $j = t + 1$. Let f be a nonzero element of $K_{ij}(\mathbf{y})$.

Theorem 5.24 implies $d(C_i) \geq i + 1 - g$ which is strictly greater than t. So $K_{ij}(\mathbf{y}) = L_j(I)$ by Lemma 6.4. So f is an error-locator function.

The function f has at most ρ_j zeros,by Lemma 5.17 since $\rho(f) \leq \rho_j$. Let $J = \{k \mid f(P_k) = 0\}$. Then J contains I, the support of $\mathbf{e}$, by Lemma 6.4. The number of elements of J is at most $\rho_j = l/2 < l + 1 - g$, since $l > 2g$. Thus $\#J < d(C_l)$ and Proposition 6.1 gives the error values.

The complexity is that of solving systems of linear equations. □

Example 6.6 This is a continuation of Example 4.17 on the Hermitian curve. Let $R = \mathbb{F}_{16}[X,Y]/(X^5 + Y^4 + Y)$, and $\rho(x) = 4$, $\rho(y) = 5$. Let us consider the 64 points on the Hermitian curve $X^5 + Y^4 + Y = 0$ over $\mathbb{F}_{16}$. As a basis for R we use the functions $x^\alpha y^\beta$, $0 \leq \alpha < 5$, $0 \leq \beta$ and then $\rho(x^\alpha y^\beta) = 4\alpha + 5\beta$.

Let φ be the evaulation map of these 64 points and let us consider the code C_{26}. This is a $[64, 38, 21]$ code. So the basic algorithm can correct $(21 - 1 - 6)/2 = 7$ errors. Let ω be a primitive element of $\mathbb{F}_{16}$ satisfying the equation $\omega^4 + \omega + 1 = 0$. We consider a seven-error pattern where the errors are located at the points $P_1 = (1, \omega)$, $P_2 = (\omega^8, \omega^3)$, $P_3 = (\omega, \omega^7)$, $P_4 = (\omega^2, \omega^3)$, $P_5 = (\omega^{11}, \omega^3)$, $P_6 = (\omega^5, \omega^3)$, $P_7 = (\omega^{14}, \omega^3)$ and with corresponding error values $e_1 = \omega^6$, $e_2 = \omega^8$, $e_3 = \omega^7$, $e_4 = \omega$, $e_5 = 1$, $e_6 = \omega^6$, and $e_7 = \omega^{10}$.

l	1	2	3	4	5	6	7	8	9	10	11	12	13
f_l	1	x	y	x^2	xy	y^2	x^3	x^2y	xy^2	y^3	x^4	x^3y	x^2y^2
ρ_l	0	4	5	8	9	10	12	13	14	15	16	17	18
s_l	ω^9	ω^{14}	0	ω^5	ω^9	ω^9	ω^7	ω^{14}	ω^{11}	ω^6	ω^2	ω^{12}	0
l	14	15	16	17	18	19	20	21	22	23	24	25	26
f_l	xy^3	y^4	x^4y	x^3y^2	x^2y^3	xy^4	y^5	x^4y^2	x^3y^3	x^2y^4	xy^5	y^6	x^4y^3
ρ_l	19	20	21	22	23	24	25	26	27	28	29	30	31
s_l	ω^4	ω^5	ω^5	ω^{12}	ω^7	ω^7	ω^6	ω^6	ω^3	ω^6	ω^4	ω^{11}	ω^{10}

In this case $t = 7$, $l = 26$ and $g = 6$ so we will use the basic algorithm $\mathcal{A}(13, 8)$. An element of $K_{13,8}$ has the form

$$\lambda_1 + \lambda_2 x + \lambda_3 y + \lambda_4 x^2 + \lambda_5 xy + \lambda_6 y^2 + \lambda_7 x^3 + \lambda_8 x^2 y$$

where the coefficients λ_i satisfy the equation

$$\begin{bmatrix} \omega^9 & \omega^{14} & 0 & \omega^5 & \omega^9 & \omega^9 & \omega^7 & \omega^{14} \\ \omega^{14} & \omega^5 & \omega^9 & \omega^7 & \omega^{14} & \omega^{11} & \omega^2 & \omega^{12} \\ 0 & \omega^9 & \omega^9 & \omega^{14} & \omega^{11} & \omega^6 & \omega^{12} & 0 \\ \omega^5 & \omega^7 & \omega^{14} & \omega^2 & \omega^{12} & 0 & \omega^5 & \omega^5 \\ \omega^9 & \omega^{14} & \omega^{11} & \omega^{12} & 0 & \omega^4 & \omega^5 & \omega^{12} \\ \omega^9 & \omega^{11} & \omega^6 & 0 & \omega^4 & \omega^5 & \omega^{12} & \omega^7 \\ \omega^7 & \omega^2 & \omega^{12} & \omega^5 & \omega^5 & \omega^{12} & 1 & \omega^5 \\ \omega^{14} & \omega^{12} & 0 & \omega^5 & \omega^{12} & \omega^7 & \omega^5 & \omega^6 \\ \omega^{11} & 0 & \omega^4 & \omega^{12} & \omega^7 & \omega^7 & \omega^6 & \omega^3 \\ \omega^6 & \omega^4 & \omega^5 & \omega^7 & \omega^7 & \omega^6 & \omega^3 & \omega^6 \\ \omega^2 & \omega^5 & \omega^5 & 1 & \omega^5 & \omega^6 & \omega^8 & \omega^{13} \\ \omega^{12} & \omega^5 & \omega^{12} & \omega^5 & \omega^6 & \omega^3 & \omega^{13} & \omega \\ 0 & \omega^{12} & \omega^7 & \omega^6 & \omega^3 & \omega^6 & \omega & \omega^{10} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \end{bmatrix} = \underline{0}$$

Here we have used that $S(x^5) = S(y^4 + y) = \omega^5 + 0 = \omega^5$ and the corresponding expressions for $S(x^6)$, $S(x^5y)$, $S(x^7)$, $S(x^6y)$, and $S(x^5y^2)$.

It can be seen that $(\lambda_1, \lambda_2, \ldots, \lambda_8) = (\omega^{11}, \omega^{13}, \omega^{13}, 0, \omega^{10}, 1, 0, 0)$ is a solution so $\omega^{11} + \omega^{13}x + \omega^{13}y + \omega^{10}xy + \omega^2 \in K_{13,8}$. The zeros of this polynomial are $P_1, \ldots, P_7$ and $(\omega^{11}, \omega^{14})$; $(\omega^{14}, \omega^{12})$; (ω^4, ω^6).

Remark 6.7 The basic algorithm can be modified by considering all basic algorithms $\mathcal{A}(i,j)$ such that $l(i,j) \leq l$ and taking the smallest j such that $K_{ij}(\mathbf{y})$ is not zero. This results in the *modified algorithm* which can correct $\lfloor (d_G(l)-1)/2 - g/4 \rfloor$ errors for codes from plane curves.

6.3 Majority voting of unknown syndromes

In this section we examine codes C_l where ρ is an order function but not necessarily a weight function.

Let $\mathbf{y}$ be a received word with error vector $\mathbf{e}$ with respect to the code C_l. If we knew the syndromes $s_i = s_i(\mathbf{e})$ for all $i \leq N$ with N as in Section 4.2, then we could solve the system of linear equations $s_i(\mathbf{x}) = s_i$ for all i, which would have the unique solution $\mathbf{x} = \mathbf{e}$. The syndromes $s_i(\mathbf{y})$ can be computed for all i, and $s_i(\mathbf{y}) = s_i(\mathbf{e})$ for all $i \leq l$. The syndrome $s_i(\mathbf{e})$ is called *known* with respect to C_l if $i \leq l$, and *unknown* if $i > l$. It will be shown how the unknown syndrome s_{l+1} can be obtained from the known ones by a *majority vote*, if the number of errors is at most $\lfloor (\nu_l - 1)/2 \rfloor$.

The matrix of syndromes $(s_{ij}(\mathbf{e}) \mid 1 \leq i,j \leq N)$ with respect to an error vector $\mathbf{e}$ was defined by:

$$s_{ij}(\mathbf{e}) = \mathbf{e} \cdot \varphi(f_i f_j)$$

in Definition 4.6. If $\mathbf{y}$ is a received word with error vector $\mathbf{e}$ with respect to the code C_l and $l(i,j) \leq l$, then $f_i f_j \in L_l$, so $s_{ij}(\mathbf{e}) = s_{ij}(\mathbf{y})$. Thus $s_{ij}(\mathbf{e})$ is a known entry of the matrix of syndromes for all i,j such that $l(i,j) \leq l$. Abbreviate $s_{ij}(\mathbf{e})$ and $s_l(\mathbf{e})$ by s_{ij} and s_l, respectively.
The set N_l was defined by

$$N_l = \{(i,j) \in \mathbb{N}^2 \mid l(i,j) = l+1 \ \}$$

in Definition 4.8. The entries of the matrix of syndromes with index $(i,j) \in N_l$ are the first unknown syndromes we encounter with respect to the code C_l. As soon as we know one s_{ij} with $(i,j) \in N_l$, we know all the other $s_{i'j'}$ with $(i',j') \in N_l$, since each one of the functions $f_i f_j$, $f_{i'} f_{j'}$ or f_{l+1} is a generator of the one dimensional vector space L_{l+1} modulo L_l. In other words, there exist $\mu_{ij}, \mu_{ijk} \in \mathbb{F}_q$ such that μ_{ij} is not zero and

$$f_i f_j = \mu_{ij} f_{l+1} + \sum_{k=1}^{l} \mu_{ijk} f_k$$

for all i, j with $l(i,j) = l+1$. Therefore

$$s_{ij} = \mu_{ij} s_{l+1} + \sum_{k=1}^{l} \mu_{ijk} s_k$$

and this relation is the same for all error vectors. Consider the matrix

$$\mathbf{S}(i,j) = (s_{i'j'}(\mathbf{e}) \mid 1 \leq i' \leq i, 1 \leq j' \leq j).$$

as was done in the previous section on the basic algorithm with the syndromes $s_{ij}(\mathbf{y})$ instead of $s_{ij}(\mathbf{e})$. If $l(i,j) = l+1$, then all entries of this matrix, except s_{ij}, are known, since $l(i',j') \leq l$ if $i' \leq i$, $j' \leq j$ and $(i',j') \neq (i,j)$.

$$\begin{pmatrix} s_{1,1} & \dots & s_{1,j-1} & s_{1,j} \\ \vdots & & \vdots & \vdots \\ s_{i-1,1} & \dots & s_{i-1,j-1} & s_{i-1,j} \\ s_{i,1} & \dots & s_{i,j-1} & ? \end{pmatrix}.$$

Remark 6.8 If $l(i,j) = l$, then $\mathbf{S}(i,j)$ is a matrix of the linear map from L_j to L_i which is used to compute the kernel $K_{ij}(\mathbf{y})$ in the basic algorithm $\mathcal{A}(i,j)$ for the code C_l. If f is a nonzero error-locator function in L_j and $f = \sum_{j'=1}^{j} \lambda_{j'} f_{j'}$, then the columns of the matrix $\mathbf{S}(i,j)$ are dependent:

$$\sum_{j'=1}^{j} s_{i'j'} \lambda_{j'} = 0 \quad \text{for all} \quad 1 \leq i' \leq i.$$

Definition 6.9 If $(i,j) \in N_l$, that is to say $l(i,j) = l+1$, and the three matrices $\mathbf{S}(i-1, j-1)$, $\mathbf{S}(i-1,j)$ and $\mathbf{S}(i, j-1)$ have equal rank, then (i,j) is called a *candidate* with respect to C_l. If (i,j) is a candidate, then there is a unique value s'_{ij} to assign to the unknown entry s_{ij} such that the matrices $\mathbf{S}(i,j)$ and $\mathbf{S}(i-1,j-1)$ have equal rank. The element s'_{ij} is called the *predicted* or *candidate value* of the unknown syndrome s_{ij}. A candidate is called *correct* or *true* when $s'_{ij} = s_{ij}$ and *incorrect* or *false* otherwise. Using the identities between the syndromes, every $(i,j) \in N_l$ gives a predicted value $s_{l+1}(i,j)$ of s_{l+1} by

$$s_{l+1}(i,j) = \frac{s'_{ij} - \sum_{k=1}^{l} \mu_{ijk} s_k}{\mu_{ij}}.$$

Denote the number of true candidates by T and the number of false candidates by F. An entry (i,j) is called a *discrepancy* if the three matrices $\mathbf{S}(i-1,j-1)$, $\mathbf{S}(i-1,j)$ and $\mathbf{S}(i,j-1)$ have equal rank and the matrices $\mathbf{S}(i,j)$ and $\mathbf{S}(i-1,j-1)$ do not have equal rank.

Remark 6.10 The discrepancies are the *pivots* if one applies *Gaussian elimination* (without interchanging rows or columns) to the matrix of syndromes. The total number of discrepancies is equal to the rank of the matrix of syndromes. The rank of the matrix of syndromes is equal to the weight of $\mathbf{e}$ by Lemma 4.7. Therefore the total number of discrepancies is equal to the number of errors.

Let $\mathbf{y}$ be a received word with error vector $\mathbf{e}$ which has at most $(\nu_l - 1)/2$ errors with respect to the code C_l. Denote the number of discrepancies in the known part of the matrix by K. A candidate is incorrect if and only if it is a discrepancy, so

$$K + F \leq \text{ total number of discrepancies } = \text{wt}(\mathbf{e}).$$

If entry (i,j) is a known discrepancy, then all entries (i,j') in the i-th row with $j' > j$, and all entries (i',j) in the j-th column with $i' > i$ are noncandidates. If $(i,j) \in N_l$ is not a candidate, then there is at least one known discrepancy in the same row or column. Thus the number of pairs $(i,j) \in N_l$ which are noncandidates is at most $2K$. The number of pairs $(i,j) \in N_l$ which are candidates is equal to $T + F$. Therefore

$$\nu_l = \text{ \# candidates } + \text{ \# noncandidates } \leq (T+F) + 2K.$$

Suppose that the number of errors is at most $(\nu_l - 1)/2$. Then

$$wt(\mathbf{e}) \leq \frac{\nu_l - 1}{2}.$$

Combining the above inequalities gives

$$F < T.$$

There may be no direct way to see whether a candidate is true or false. But a predicted value s'_{ij} of the syndrome $s_{i,j}$ is assigned to every candidate, and this gives a predicted value or vote $s_{l+1}(i,j)$ for s_{l+1} by Definition 6.9. All T true candidates yield the same, correct, value for s_{l+1}. Thus a proof of the following proposition has been given.

Proposition 6.11 *If the number of errors of a received word with respect to the code C_l is at most $(\nu_l - 1)/2$, then the majority of the candidates vote for the correct value of s_{l+1}.*

Hence by recursion all the unknown sydromes with respect to the code C_l can be obtained if the number of errors is at most $\lfloor (d_\varphi(l) - 1)/2 \rfloor$, since $\nu_m \geq d_\varphi(l)$ if $m \geq l$ and $C_{m+1} \neq C_m$.

From this the error vector is obtained. The complexity of the algorithm is at most the complexity of solving a system of n linear equations in n unknowns, which is at most $\mathcal{O}(n^3)$ for $n \to \infty$. Thus the proof of the following theorem has been given.

Theorem 6.12 *$\lfloor d_\varphi(l) - 1)/2 \rfloor$ errors are corrected for the code C_l by majority voting for unknown syndromes with complexity $\mathcal{O}(n^3)$.*

Remark 6.13 If ρ is a weight function with g gaps, then it is not neccessary to compute all unknown syndromes. One could stop as soon as one has the unknown syndromes $s_{l+1}, \ldots, s_{l+g}$ and apply the basic algorithm to the code C_{l+g}.

Example 6.14 The code C_{20} is a $[64, 44, 15]$-code. Let us consider the same 7-error pattern as in Example 6.6.

The syndrome matrix is (in powers of ω,with * indicating a zero)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	9	14	*	5	9	9	7	14	11	6	2	12	*	4	5	5	12	7	7	6	×
2	14	5	9	7	14	11	2	12	*	4	5	5	12	7	7	5	×				
3	*	9	9	14	11	6	12	*	4	5	5	12	7	7	6	×					
4	5	7	14	2	12	*	5	5	12	7	0	5	×								
5	9	14	11	12	*	4	5	12	7	7	5	×									
6	9	11	6	*	4	5	12	7	7	6	×										
7	7	2	12	5	5	12	0	5	×												
8	14	12	*	5	12	7	5	×													
9	11	*	4	12	7	7	×														
10	6	4	5	7	7	6															
11	2	5	5	0	5	×															
12	12	5	12	5	×																
13	*	12	7	×																	
14	4	7	7																		
15	5	7	6																		
16	5	5	×																		
17	12	×																			
18	7																				
19	7																				
20	6																				
21	×																				

Here × denotes the syndrome corresponding to f_{21}, which is the first unknown.

Using row operations one gets

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	9	14	*	5	9	9	7	14	11	6	2	12	*	4	5	5	12	7	7	6	×
2	*	8	9	6	*	10	7	6	1	13	13	1	12	0	6	0	×				
3	*	*	13	1	11	1	9	7	10	12	12	7	5	14	10	×					
4	*	*	*	5	*	*	10	8	*	*	4	13	×								
5	*	*	*	*	*	*	5	*	*	*	8	×									
6	*	*	*	*	*	*	*	*	*	*	×										
7	*	*	*	*	5	0	*	10	×												
8	*	*	*	*	*	*	*	×													
9	*	*	*	*	*	*	×														
10	*	*	*	*	*	*															
11	*	*	*	*	*	×															
12	*	*	*	*	×																
13	*	*	*	×																	
14	*	*	*																		
15	*	*	*																		
16	*	*	×																		
17	*	×																			
18	*																				
19	*																				
20	*																				
21	×																				

From this it is seen that the positions $(6, 11)$, $(8, 8)$, and $(11, 6)$ are candidate positions and by keeping track of the performed row operations one sees also that the candidate values in all three cases are ω^6, so this is the correct syndrome $S(f_{21})$. In the same manner one finds $S(f_{22}) = \omega^3$, $S(f_{23}) = \omega^6$, $S(f_{24}) = \omega^4$, $S(f_{25}) = \omega^{11}$, $S(f_{26}) = \omega^{10}$, which corresponds to the results of Example 6.6 and we can now apply the basic algorithm for the code C_{26} as before.

An algorithm will be given in the next section that computes the unknown syndromes from the known ones, and a basis for the error-locator functions.

6.4 Notes

The chapter on complexity issues [7] in coding theory is used as a reference on this topic.

The basic and modified algorithm were presented for the class of codes on plane curves by [50]. They are generalizations of the decoding algorithm for RS codes of [5] and [79]. See also [10]. The general case was treated by [93]. It was discovered independently by [56]. Another decoding algorithm is presented in [80]. It is described and proved in [81]. The equivalence with the modified algorithm was shown by [20, 21]. The error-correcting capacity of the modified algorithm was determined in [17].

By applying several basic algorithms in parallel, one can decode up to half the minimum distance as was proved in [72, 102]. Part of this approach is not constructive.

The idea of majority voting for unknown syndromes is from [24]. See also [18, 19] with the notion of *majority coset decoding*. Another explicit and constructive algorithm which decodes up to half the designed minimum distance was given by [22].

The *Fundamental Iterative Algorithm* [28] and the *Modified Fundamental Iterative Algorithm* [24] are generalizations of Gaussian elimination for a partial matrix and to get the unknown syndromes, respectively.

The survey paper [47] contains much more details and history of the decoding algorithms of AG codes.

7 Fast decoding up to half the order bound

In this section we will present an efficient computational procedure which implements the basic and the modified algorithms of Section 6.2 and an extension which includes the majority voting presented in Section 6.3.

7.1 Determination of the unknown syndromes

We recall the definition of the codes C_l. Let R be an $\mathbb{F}_q$-algebra with an order function ρ. Let φ be a surjective morphism $\varphi : R \rightarrow \mathbb{F}_q^n$ of $\mathbb{F}_q$-algebras. Fix a basis $(f_i | i \in \mathbb{N})$ of R such that $\rho(f_i) < \rho(f_{i+1})$. The codes C_l were defined as

$$C_l = \{ \ \mathbf{c} \in \mathbb{F}_q^n \mid \mathbf{c} \cdot \varphi(f) = 0 \text{ for all } f \text{ with } \rho(f) \leq \rho_l \ \}.$$

The number $l(i, j)$ was defined as the unique l such that $\rho(f_i f_j) = \rho_l$. This can be used to define the binary operation $\oplus$ on $\mathbb{N}$ by $i \oplus j = l(i, j)$. The operation is associative $(i \oplus j) \oplus k = i \oplus (j \oplus k)$, which follows from the fact that multiplication in R is associative and properties of the order function.

We define a partial order $\leq_p$ on $\mathbb{N}$ by $i \leq_p j$ if there exists a k such that $i \oplus k = j$. This is indeed a partial order, since reflexivity and transitivity are immediate and the antisymmetry follows from Lemma 3.9. The k above is unique and we denote that by $j \ominus i$.

Example 7.1 Let $R = \mathbb{F}_q[X_1, \ldots, X_m]$. Let $\prec_D$ be the graded lexicographic order and ρ the corresponding order function. Let $f_i = X^\alpha$ and $f_j = X^\beta$ be the i-th and j-th monomial with respect to the order $\prec_D$. Then $i \leq_p j$ if and only if $\alpha_t \leq \beta_t$ for all t, and $f_{i \oplus j} = X^{\alpha+\beta}$. If $j \leq_p i$, then $f_{i \ominus j} = X^{\alpha - \beta}$.

Example 7.2 Let R be an $\mathbb{F}_q$-algebra with a weight function ρ. Let g be the number of gaps and c the conductor of the associated semigroup. Then $\rho_i = i + g - 1$ if $i > c - g$ by Lemma 5.6. If $i, j > c - g$, then $\rho_{i \oplus j} = \rho_i + \rho_j = i + j + 2g - 2$, so $i \oplus j = i + j + g - 1$.

For an element $\mathbf{y} \in \mathbb{F}_q^n$ we define

$$S_{\mathbf{y}}(f) = \mathbf{y} \cdot \varphi(f),$$

so in particular we have that $S_{\mathbf{y}}(f_i) = s_i(\mathbf{y})$ and $S_{\mathbf{y}}(f_i f_j) = s_{ij}(\mathbf{y})$ are the syndromes of Definition 4.6. In the decoding situation, $\mathbf{y} = \mathbf{c} + \mathbf{e}$ is a received word, $\mathbf{c} \in C_l$ and $\mathbf{e}$ is the error with respect to C_l. The syndromes $S_{\mathbf{e}}(f)$ are equal to $S_{\mathbf{y}}(f)$ if $\rho(f) \leq \rho_l$ and can be calculated directly from $\mathbf{y}$. The algorithm takes as input the known syndromes in increasing order and gives as output the syndromes $S_{\mathbf{e}}(f)$ where $\rho(f) > \rho_l$ provided that

$$\mathrm{wt}(\mathbf{e}) \leq \lfloor (d(l) - 1)/2 \rfloor.$$

In the following let $\mathbf{e}$ be fixed. We will omit the subscript $\mathbf{e}$ in the syndromes. So $S(f) = S_{\mathbf{e}}(f)$.

Definition 7.3 For $f \in R$ with $\rho(f) = \rho_j$ we define the *span* and *fail* of f by

$$\mathrm{span}(f) = i \text{ if } S(ff_i) \neq 0 \text{ but } S(ff_k) = 0 \text{ for all } k < i,$$

$$\mathrm{fail}(f) = \mathrm{span}(f) \oplus j.$$

This means that i is the largest number such that f is in the kernel $K_{i-1,j}$ as defined in 6.2. If $\text{span}(f) = i$ and $\rho(g) = \rho_k$, then $S(fg) = 0$ if $k < i$, and $S(fg) \neq 0$ if $k = i$. If $\text{fail}(f) = k$, then ρ_k is the smallest order for which some multiple of f has that order and a nonzero syndrome.

Before presenting the algorithm, we will prove a series of lemmas which eventually will prove that the algorithm works.

Lemma 7.4 *Let $f \in R$, $\rho(f) = \rho_k$ and $\text{span}(f) = i$. If $g \in R$ and $\rho(g) = \rho_i$, then $\text{span}(g) \leq k$. More generally if $j \leq_p i$ and $\rho(g) = \rho_{i \ominus j}$, then $\text{span}(g) \leq k \oplus j$ and $\text{fail}(g) \leq \text{fail}(f)$.*

Proof. If $\rho(g) = \rho_{i\ominus j}$, then $\rho(gf_j) = \rho_i$. Since $\text{span}(f) = i$, $S(f(gf_j)) \neq 0$. So $S(g(ff_j)) \neq 0$. Now $\rho(ff_j) = \rho_{k\oplus j}$. Therefore $\text{span}(g) \leq k \oplus j$. From this we get

$$\text{fail}(g) \leq k \oplus j \oplus (i \ominus j) = k \oplus i = \text{fail}(f).$$

□

Definition 7.5 For $l \in \mathbb{N}$ let

$$\Sigma_l = \{i \mid \text{there exists an } f \in R \text{ such that } \rho(f) = \rho_i \text{ and } \text{fail}(f) > l\}$$

and let σ_l be the set of minimal elements of Σ_l with respect to $\leq_p$. Let

$$\Delta_l = \{\text{span}(f) \mid f \in R, \text{fail}(f) \leq l\}$$

and let δ_l be the set of maximal elements of Δ_l with respect to $\leq_p$.

Lemma 7.6

$$\Sigma_l \cap \Delta_l = \emptyset$$

Proof. Let $i \in \Delta_l$ and let $f \in R$ have $\text{span}(f) = i$ and $\text{fail}(f) \leq l$. We want to show that $i \notin \Sigma_l$. But if $g \in R$ with $\rho(g) = \rho_i$, then it follows from Lemma 7.4 that $\text{fail}(g) \leq \text{fail}(f) \leq l$ and therefore $i \notin \Sigma_l$. □

Lemma 7.6 shows that the sets Σ_l and Δ_l are disjoint. We will prove that

$$\Sigma_l \cup \Delta_l = \mathbb{N}.$$

This will be a result of an iterative procedure which also shows how the set $\text{Info}_l = \{\sigma_l, \delta_l, F_l, G_l\}$ can be computed from Info_{l-1}, where F_l and G_l are mappings

$$F_l : \sigma_l \to R \;\text{ and }\; G_l : \delta_l \to R,$$

such that $F_l(i) = f$ is a choice of an element in $f \in R$ with $\rho(f) = \rho_i$ and $\text{fail}(f) > l$, and $G_l(i) = g$ is a choice of an element in $g \in R$ with $\text{span}(g) = i$ and $\text{fail}(g) \leq l$.

That this is indeed the case rests on the following lemmas.

Lemma 7.7 *Let $f \in R$, $\rho(f) = \rho_k$ and* $\text{span}(f) = i$. *Let $g \in R$ with $\rho(g) = j$. If $j \leq_p i$, then* $\text{span}(fg) = i \ominus j$ *and* $\text{fail}(fg) = \text{fail}(f) = k \oplus i$. *If $j \not\leq_p i$, then* $\text{span}(fg) > m$ *for all m such that $j \oplus m < i$.*

Proof. Now $S((fg)f_l) = S(f(gf_l))$. So if $\rho(gf_l) < \rho_i$ we have $S(f(gf_l)) = 0$ and if $\rho(gf_l) = \rho_i$ we have $S(f(gf_l)) \neq 0$. Therefore if $j \leq_p i$, then $\text{span}(fg) = i \ominus j$ and $\text{fail}(fg) = (i \ominus j) \oplus (k \oplus j) = k \oplus i = \text{fail}(f)$. Furthermore if $j \not\leq_p i$, then $\text{span}(fg) > m$ for all m such that $j \oplus m < i$. □

Lemma 7.8 *Let $j \leq_p i$.*
(1) If $j \in \Sigma_l$, then $i \in \Sigma_l$.
(2) If $i \in \Delta_l$, then $j \in \Delta_l$.

Proof. (1) Let $j \in \Sigma_l$. Then there exists an $f \in R$ such that $\rho(f) = \rho_j$ and $\text{fail}(f) > l$. If $j \leq_p i$, then $\rho(ff_{i \ominus j}) = \rho_i$ and $\text{fail}(ff_{i \ominus j}) > l$. So $i \in \Sigma_l$.

The proof of (2) is similar. □

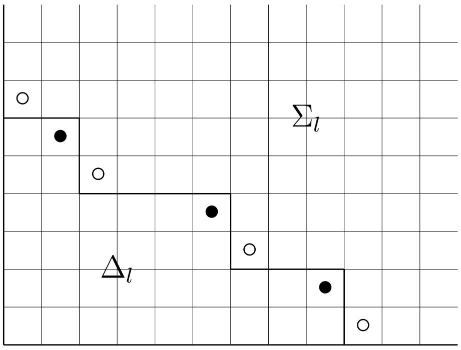
So $(\Sigma_l | l \in \mathbb{N})$ is a sequence of subsets of $\mathbb{N}$ which is decreasing with respect to inclusion, and σ_l is the set of minimal elements of Σ_l with respect to $\leq_p$. So

$$\Sigma_l = \{i \in \mathbb{N} | \text{ there exists a } j \in \sigma_l \text{ such that } j \leq_p i\}.$$

Similarly, $(\Delta_l | l \in \mathbb{N})$ is an increasing sequence of subsets of $\mathbb{N}$ and δ_l is the set of maximal elements of Δ_l. So

$$\Delta_l = \{j \in \mathbb{N} | \text{ there exists an } i \in \delta_l \text{ such that } j \leq_p i\}.$$

Example 7.9 Below a picture is given for the graded lexicographic order $\prec$ on the monomials in two variables. Here $(\alpha, \beta) \in \mathbb{N}_0^2$ is mapped to $i \in \mathbb{N}$, where $X^\alpha Y^\beta$ is the i-th monomial with respect to $\prec$. In this way the subsets Δ_l and Σ_l of $\mathbb{N}$ can be identified with subsets of $\mathbb{N}_0^2$.



The bullets • denote the positions of the elements of δ_l, which are the maximal elements of Δ_l. The circles ∘ denote the positions of the elements of σ_l, which are the minimal elements of Σ_l.

Lemma 7.10 *Let $f \in R$, $\rho(f) = \rho_k$ and* $\mathrm{span}(f) = i$. *If $\rho(g) = \rho_j$ and $j \not\leq_p i$, then $\rho(fg) = \rho_{k \oplus j}$ and* $\mathrm{fail}(fg) > k \oplus i$.

Proof. That $\rho(fg) = \rho_{k \oplus j}$ follows from the definition of $\oplus$. Let $m = \mathrm{span}(fg)$. If $\mathrm{fail}(fg) \leq k \oplus i$, then $m \oplus k \oplus j \leq k \oplus i$. So $m \oplus j \leq i$. But $j \not\leq_p i$. Hence $j \oplus m < i$. Therefore $\mathrm{span}(fg) > m$ by Lemma 7.7, which is a contradiction. □

Lemma 7.11 *Let $f, f' \in R$, $\rho(f) = \rho_k$, $\rho(f') = \rho_{k'}$,* $\mathrm{span}(f) = l$ *and* $\mathrm{span}(f') = l'$. *Let $g, g' \in R$, $\rho(g) = \rho_i$ and $\rho(g') = \rho_{i'}$. Suppose $k \oplus l > k' \oplus l'$. If $i \leq_p l$, $i' \leq_p l'$ and $l \ominus i = l' \ominus i'$, then there exists a $\mu \in \mathbb{F}_q^*$ such that $h = fg + \mu f'g'$ satisfies $\rho(h) = \rho_{k \oplus i}$,* $\mathrm{span}(h) > l \ominus i$ *and* $\mathrm{fail}(h) > k \oplus l$.

Proof. We have $\rho(fg) = \rho_{k \oplus i}$ and $k \oplus i = k \oplus l \ominus l' \oplus i' > k' \oplus l' \ominus l' \oplus i' = k' \oplus i'$. Furthermore $\rho_{k' \oplus i'} = \rho(\mu f'g')$, so $\rho(h) = \rho_{k \oplus i}$ for all $\mu \in \mathbb{F}_q^*$.

Let $h' \in R$ have $\rho(h') = \rho_{l \ominus i} = \rho_{l' \ominus i'}$ and let $\lambda = S(fgh')$ and $\lambda' = S(f'g'h')$. Here both λ and λ' are nonzero since $\mathrm{span}(f) = l = \rho(gh')$ and $\mathrm{span}(f') = l' = \rho(g'h')$. Letting $\mu = -\lambda/\lambda'$ gives the result. □

We describe an algorithm to obtain Info_l from Info_{l-1} and show by induction that $\Sigma_l \cup \Delta_l = \mathbb{N}$. As initial step we let $\delta_0 = \emptyset$, $\sigma_0 = \{1\}$, $F_0(1) = 1$ and $G_0 = \emptyset$ and we put $\rho_0 = -\infty$. Now suppose we have proved the statement up to $l - 1$. We will prove it for l.

Definition 7.12 Define

$$\delta'_l = \{\mathrm{span}(f) \mid f \in \mathrm{Im}(F_{l-1}) \text{ and } \mathrm{fail}(f) = l\}.$$

$$\Delta'_l = \Delta_{l-1} \cup \{i \ominus j \mid i \in \delta'_l, j \leq_p i\}.$$

Let σ'_l be the set of minimal elements of $\mathbb{N} \setminus \Delta'_l$ with respect to $\leq_p$.

Lemma 7.13

$$\Delta'_l \subseteq \Delta_l.$$

Proof. It was already remarked that $\Delta_{l-1} \subseteq \Delta_l$.

Let $i \in \delta'_l$. Then there exists a $k \in \sigma_{l-1}$ such that $\mathrm{span}(f) = i$ and $\mathrm{fail}(f) = l$, where $f = F_{l-1}(k)$. Let $j \leq_p i$. Then $\mathrm{span}(ff_j) = i \ominus j$ and $\mathrm{fail}(ff_j) = \mathrm{fail}(f)$ by Lemma 7.7. Hence $i \ominus j \in \Delta_l$. □

We will construct for each $i \in \sigma'_l$ a function of order ρ_i and fail $> l$ thus proving that $\sigma'_l \subseteq \Sigma_l$. Therefore $\mathbb{N} \setminus \Delta'_l \subseteq \Sigma_l$ by Lemma 7.8. But $\Delta'_l \subseteq \Delta_l$ by Lemma 7.13. So $\mathbb{N} \setminus \Delta_l \subseteq \Sigma_l$. Lemma 7.6 says that $\Delta_l \cap \Sigma_l = \emptyset$. Hence $\Sigma_l = \mathbb{N} \setminus \Delta_l$. From this of course also follows that $\Delta_l = \Delta'_l$ and $\sigma_l = \sigma'_l$. So the set of maximal elements of Δ'_l is δ_l. Hence δ_l is contained in $\delta_{l-1} \cup \delta'_l$. Possibly there exists a $j \in \delta_{l-1}$ and a $j' \in \delta'_l$ such that $j \leq_p j'$. Therefore δ_l is the disjoint union of δ'_l and $\{j \in \delta_{l-1} \mid j \not\leq_p j' \text{ for all } j' \in \delta'_l\}$. Furthermore F_l and G_l will be constructed from F_{l-1} and G_{l-1}, respectively.

Let $i \in \sigma'_l$. Then $i \in (\mathbb{N} \setminus \Delta'_l) \subseteq (\mathbb{N} \setminus \Delta_{l-1})$ which is equal to Σ_{l-1} by the induction hypthesis. So $i = k \oplus j$ for some $k \in \sigma_{l-1}$. Let $f = F_{l-1}(k)$. Let $m = \mathrm{span}(f)$. Then $\mathrm{fail}(f) = m \oplus k$. If $\mathrm{fail}(f) = l$, then $l \ominus i \notin \Sigma_{l-1}$. Otherwise $l \ominus i = k' \oplus j'$ for some $k' \in \sigma_{l-1}$. We can then apply Lemma 7.4 with f and $g = F_{l-1}(k')$ since $\rho(g) = \rho_{k'}$ and $k' = l \ominus i \ominus j' = l \ominus (k \oplus j) \ominus j' = m \ominus (j \oplus j')$ and therefore get that $\mathrm{fail}(g) \leq k \oplus m = l$ and in fact $\mathrm{fail}(g) = l$ since $\mathrm{fail}(g) \geq l$ by assumption. But this implies $j' \oplus i \in \delta'_l$ and therefore $i \in \Delta'_l$ contradicting the fact that $i \in \sigma'_l$.

We will consider two cases for i.

(1) Suppose $i \in \sigma_{l-1}$. So $i = k$. Let $f = F_{l-1}(k)$. If $\mathrm{fail}(f) > l$ we are done and let $F_l(k) = f$. If $\mathrm{fail}(f) = l$, then by the above remark we have $l \ominus i \notin \Sigma_{l-1}$. So $l \ominus i \in \Delta_{l-1}$, by the induction hypothesis, and we can find elements $l' \in \delta_{l-1}$ and $i' \in \mathbb{N}$ such that $l \ominus i = l' \ominus i'$. Let $f' = G_{l-1}(l')$. Let $g = 1$ and $g' = f_{i'}$. Then applying Lemma 7.11 to f, f', g and g' gives a new function of order ρ_k and fail strictly larger than l.

(2) Suppose $i \notin \sigma_{l-1}$. Then $i = k \oplus j$ for $k \in \sigma_{l-1}$ and $j > 1$. Let $g = f_j$. Now $f = F_{l-1}(k)$ must have $\mathrm{fail}(f) = l$ because $k \in \Delta_l$. By the claim above, either $i \not\leq_p l$ or $l \ominus i \in \Delta_{l-1}$. In the first case fg has order ρ_i and $\mathrm{fail}(fg) > l$ by Lemma 7.10. In the second case $l \ominus i = l' \ominus i'$ for some $l' \in \delta_{l-1}$. Let $f' = G_{l-1}(l')$. Let $g = f_j$ and $g' = f_{i'}$. Then applying Lemma 7.11 to f, f', g and g' gives a new function of order $\rho_{k \oplus j}$ and fail strictly larger than l.

For each $i \in \sigma'_l$, we have constructed a function of order ρ_i and fail strictly larger than l. This proves the claim. We have also produced the sets σ_l and δ_l and the function F_l.

For the function G_l we note that each element i of δ_l is either an element of δ_{l-1} or δ'_l. In the first case, $G_l(i) = G_{l-1}(i)$ and in the second case $G_l(i) = F_{l-1}(l \ominus i)$.

We can now formulate the following algorithm.

Algorithm 7.14

```
Initialization: $\delta_0 = \emptyset, \sigma_0 = \{1\}, F_0(1) = 1, G_0 = \emptyset$.
Given $\mathrm{Info}_{l-1} = \{\sigma_{l-1}, \delta_{l-1}, F_{l-1}, G_{l-1}\}$
(0)  Let $\delta'_l = \{\mathrm{span}(f) \mid f \in \mathrm{Im}(F_{l-1}) \text{ and } \mathrm{fail}(f) = l\}$
     Let $\Delta_l = \Delta_{l-1} \cup \{i \ominus j \mid i \in \delta'_l, j \leq_p i\}$
     Let $\sigma_l$ be the set of minimal elements of $\mathbb{N} \setminus \Delta_l$
(1)  for each $i \in \sigma_{l-1}$
         let $f = F_{l-1}(i)$
         if $i \not\leq_p l$ or $S(f f_{l \ominus i}) = 0$, then
             $i \in \sigma_l$ and $F_l(i) = f$
         else
             a)  if $l \ominus i = l' \ominus i'$ for $l' \in \delta_{l-1}$ and $i' \in \mathbb{N}$, then
                     $i \in \sigma_l$ and $f' = G_{l-1}(l')$, $F_l(k) = fg + \mu f' g'$
                     (as in Lemma 7.11 with $g = 1$ and $g' = f_{i'}$)
                 else
             b)  $l \ominus i \in \delta'_l$ and $G_l(l \ominus i) = f$
(2)  for each $i \in \sigma_l \setminus \sigma_{l-1}$
         $i = k \oplus j$, where $k \in \sigma_{l-1}$ and $j > 1$, let $f = F_{l-1}(k)$
             if $i \not\leq_p l$, then $F_l(i) = f_j f$
             else
                 $l \ominus i = l' \ominus i'$ for $l' \in \delta_{l-1}$ and
```

$f' = G_{l-1}(l')$, $F_l(i) = fg + \mu f'g'$
(as in Lemma 7.11 with $g = f_j$ and $g' = f_{i'}$)

We have proved the following theorem.

Theorem 7.15 *For all $l \in \mathbb{N}$ the two sets Σ_l and Δ_l partition $\mathbb{N}$ and Algorithm 7.14 gives as output* Info_l *with* Info_{l-1} *as input. .*

Example 7.16 This is a continuation of Example 6.14 on the Hermitian curve. Consider the code C_{27} with parameters $[64, 44, 15]$. Take the same error-vector with 7 errors as in Example 6.6. Consider the following 20 known syndromes.

l	1	2	3	4	5	6	7	8	9	10
s_l	ω^9	ω^{14}	0	ω^5	ω^9	ω^9	ω^7	ω^{14}	ω^{11}	ω^6
l	11	12	13	14	15	16	17	18	19	20
s_l	ω^2	ω^{12}	0	ω^4	ω^5	ω^5	ω^{12}	ω^7	ω^7	ω^6

In the following table, l is in the first column and f_l is in the second. The third, fourth and fifth column are split into two rows for every l, corresponding to σ_l and δ_l, and the fifth column gives the functions that are the image under F_l and G_l, respectively, of elements in the fourth column.

0		σ_0	1	1
		δ_0		
1	1	σ_1	2	x
			3	y
		δ_1	1	1
2	x	σ_2	2	$x + \omega^5$
			3	y
		δ_2	1	1
3	y	σ_3	2	$x + \omega^5$
			3	y
		δ_3	1	1
4	x^2	σ_4	3	y
			4	$x(x + \omega^5) + \beta - 1 = x^2 + \omega^5 x + \omega^{14}$
		δ_4	2	$x + \omega^5$
5	xy	σ_5	3	$y + \omega x + \omega^6$
			4	$x^2 + \omega^5 x + \omega^{14}$
		δ_5	2	$x + \omega^5$
6	y^2	σ_6	4	$x^2 + \omega^5 x + \omega^{14}$
			5	$x(y + \omega x + \omega^6) = xy + \omega x^2 + \omega^6 x$
			6	$y(y + \omega x + \omega^6) + \omega^5 x(x + \omega^5)$
				$= y^2 + \omega xy + \omega^5 x^2 + \omega^6 y + \omega^{10} x$
		δ_6	2	$x + \omega^5$
			3	$y + \omega x + \omega^6$
7	x^3	σ_7	4	$(x^2 + \omega^5 x + \omega^{14}) + \omega^7(x + \omega^5) = x^2 + \omega^{13} x + \omega^5$
			5	$xy + \omega x^2 + \omega^6 x$
			6	$y^2 + \omega xy + \omega^5 x^2 + \omega^6 y + \omega^{10} x$
		δ_7	2	$x + \omega^5$
			3	$y + \omega x + \omega^6$
8	$x^2 y$	σ_8	4	$x^2 + \omega^{13} x + \omega^5 + \omega^3(y + \omega x + \omega^6) = x^2 + \omega^3 y + \omega^{11} x + \omega^6$
			5	$xy + \omega x^2 + \omega^6 x + \omega^8(x + \omega^5) = xy + \omega x^2 + \omega^{14} x + \omega^{13}$
			6	$y^2 + \omega xy + \omega^5 x^2 + \omega^6 y + \omega^{10} x$
		δ_8	2	$x + \omega^5$
			3	$y + \omega x + \omega^6$
9	xy^2	σ_9	4	$x^2 + \omega^3 y + \omega^{11} x + \omega^6$
			5	$xy + \omega x^2 + \omega^{14} x + \omega^{13} + \omega^{11}(y + \omega x + \omega^6)$
				$= xy + \omega x^2 + \omega^{11} y + \omega^5 x + \omega^{14}$
			6	$y^2 + \omega xy + \omega^5 x^2 + \omega^6 y + \omega^{10} x$
		δ_9	2	$x + \omega^5$
			3	$y + \omega x + \omega^6$

10	y^3	σ_{10}	4	$x^2+\omega^3y+\omega^{11}x+\omega^6$
			5	$xy+\omega x^2+\omega^{11}y+\omega^5x+\omega^{14}$
			6	$y^2+\omega xy+\omega^5x^2+\omega^6y+\omega^{10}x+\omega^3(y+\omega x+\omega^6)$
				$=y^2+\omega xy+\omega^5x^2+\omega^2y+\omega^2x+\omega^9$
		δ_{10}	2	$x+\omega^5$
			3	$y+\omega x+\omega^6$
11	x^4	σ_{11}	5	$xy+\omega x^2+\omega^{11}y+\omega^5x+\omega^{14}$
			6	$y^2+\omega xy+\omega^5x^2+\omega^2y+\omega^2x+\omega^9$
			7	$x(x^2+\omega^3y+\omega^{11}x+\omega^6)+\omega^{12}(x+\omega^5)$
				$=x^3+\omega^3xy+\omega^{11}x^2+\omega^4x+\omega^2$
		δ_{11}	4	$x^2+\omega^3y+\omega^{11}x+\omega^6$
			3	$y+\omega x+\omega^6$
12	x^3y	σ_{12}	5	$xy+\omega x^2+\omega^{11}y+\omega^5x+\omega^{14}+\omega(x^2+\omega^3y$
				$+\omega^{11}x+\omega^6)=xy+\omega^{13}y+\omega^{14}x+\omega$
			6	$y^2+\omega xy+\omega^5x^2+\omega^2y+\omega^2x+\omega^9$
			7	$x^3+\omega^3xy+\omega^{11}x^2+\omega^4x+\omega^2+\omega^8(y+\omega x+\omega^6)$
				$=x^3+\omega^3xy+\omega^{11}x^2+\omega^8y+\omega^{14}x+\omega^{13}$
		δ_{12}	3	$y+\omega x+\omega^6$
			4	$x^2+\omega^3y+\omega^{11}x+\omega^6$
13	x^2y^2	σ_{13}	5	$xy+\omega^{13}y+\omega^{14}x+\omega$
			6	$y^2+\omega xy+\omega^5x^2+\omega^2y+\omega^2x+\omega^9+\omega^5(x^2+\omega^3y$
				$+\omega^{11}x+\omega^6)=y^2+\omega xy+y+\omega^5x+\omega^2$
			7	$x^3+\omega^3xy+\omega^{11}x^2+\omega^8y+\omega^{14}x+\omega^{13}$
		δ_{13}	3	$y+\omega x+\omega^6$
			4	$x^2+\omega^3y+\omega^{11}x+\omega^6$
14	xy^3	σ_{14}	5	$xy+\omega^{13}y+\omega^{14}x+\omega$
			6	$y^2+\omega xy+y+\omega^5x+\omega^2$
			7	$x^3+\omega^3xy+\omega^{11}x^2+\omega^8y+\omega^{14}x+\omega^{13}$
		δ_{14}	3	$y+\omega x+\omega^6$
			4	$x^2+\omega^3y+\omega^{11}x+\omega^6$
15	y^4	σ_{15}	5	$xy+\omega^{13}y+\omega^{14}x+\omega$
			6	$y^2+\omega xy+y+\omega^5x+\omega^2$
			7	$x^3+\omega^3xy+\omega^{11}x^2+\omega^8y+\omega^{14}x+\omega^{13}+\omega^3(x^2+\omega^3y$
				$+\omega^{11}x+\omega^6)=x^3+\omega^3xy+\omega^5x^2+\omega^{14}y+\omega^{10}$
		δ_{15}	3	$y+\omega x+\omega^6$
			4	$x^2+\omega^3y+\omega^{11}x+\omega^6$

16	x^4y	σ_{16}	6	$y^2 + \omega xy + y + \omega^5 x + \omega^2$
			8	$x(xy + \omega^{13}y + \omega^{14}x + \omega) + 1 \cdot (x^2 + \omega^3 y + \omega^{11}x + \omega^6)$
				$= x^2y + \omega^{13}xy + \omega^3x^2 + \omega^3y + \omega^6x + \omega^6$
			11	$x(x^3 + \omega^3xy + \omega^5x^2 + \omega^{14}y + \omega^{10}) + \omega^7(y + \omega x + \omega^6)$
				$= x^4 + \omega^3x^2y + \omega^5x^3 + \omega^{14}xy + \omega x + \omega^{13} + \omega^7y$
		δ_{16}	7	$xy + \omega^{13}y + \omega^{14}x + \omega$
			5	$x^3 + \omega^3xy + \omega^5x^2 + \omega^{14}y + \omega^{10}$
17	x^3y^2	σ_{17}	6	$y^2 + \omega xy + y + \omega^5x + \omega^2 + \omega^8(xy + \omega^{13}y + \omega^{14}x + \omega)$
				$= y^2 + \omega^{10}xy + \omega^{13}y + \omega^{13}x + \omega^{11}$
			8	$x^2y + \omega^{13}xy + \omega^3x^2 + \omega^3y + \omega^6x + \omega^6$
			11	$x^4 + \omega^3x^2y + \omega^7y + \omega^5x^3 + \omega^{14}xy + \omega x + \omega^{13}$
		δ_{17}	7	$xy + \omega^{13}y + \omega^{14}x + \omega$
			5	$x^3 + \omega^3xy + \omega^5x^2 + \omega^{14}y + \omega^{10}$
18	x^2y^3	σ_{18}	6	$y^2 + \omega^{10}xy + \omega^{13}y + \omega^{13}x + \omega^{11}$
			8	$x^2y + \omega^{13}xy + \omega^3x^2 + \omega^3y + \omega^6x + \omega^6$
			11	$x^4 + \omega^3x^2y + \omega^5x^3 + \omega^{14}xy + \omega x + \omega^{13} + \omega^7y$
		δ_{18}	7	$x^3 + \omega^3xy + \omega^5x^2 + \omega^{14}y + \omega^{10}$
			5	$x^3 + \omega^3xy + \omega^5x^2 + \omega^{14}y + \omega^{10}$
19	xy^4	σ_{19}	6	$y^2 + \omega^{10}xy + \omega^{13}y + \omega^{13}x + \omega^{11}$
			8	$x^2y + \omega^{13}xy + \omega^3x^2 + \omega^3y + \omega^6x + \omega^6x^4$
				$+\omega^3x^2y + \omega^5x^3 + \omega^{14}xy + \omega^7y + \omega^{13}$
			11	$\omega x(xy + \omega^{13}y + \omega^{14}x + \omega)$
				$= x^4 + \omega^5x^3 + \omega^9x^2y + x^2 + \omega^7y + \omega^5x + \omega^{13}$
		δ_{19}	7	$xy + \omega^{13}y + \omega^{14}x + \omega$
			5	$x^3 + \omega^3xy + \omega^5x^2 + \omega^{14}y + \omega^{10}$
20	y^5	σ_{20}	6	$y^2 + \omega^{10}xy + \omega^{13}y + \omega^{13}x + \omega^{11}$
			8	$x^2y + \omega^{13}xy + \omega^3x^2 + \omega^3y + \omega^6x + \omega^6$
				$+\omega^{11}(xy + \omega^{13}y + \omega^{14}x + \omega)$
				$= x^2y + \omega^4xy + \omega^3x^2 + \omega y + \omega^7x + \omega^4$
			11	$x^4 + \omega^5x^3 + \omega^9x^2y + x^2 + \omega^7y + \omega^5x + \omega^{13}$
				$+\omega^{11}(x^3 + \omega^3xy + \omega^5x^2 + \omega^{14}y + \omega^{10})$
				$= x^4 + \omega^3x^3 + \omega^9x^2y + \omega^4x^2 + \omega^{14}xy + \omega^6y + \omega^5x + 1$
		δ_{20}	7	$xy + \omega^{13}y + \omega^{14}x + \omega$
			5	$x^3 + \omega^3xy + \omega^5x^2 + \omega^{14}y + \omega^{10}$

Running the algorithm from 0 to l will produce F_l and σ_l and it follows from the discussion that if s is a minimal element of σ_l and $F_l(s) = f$ then f is an element of $K_{l \ominus s, s}$ and f is an error locator function if $d(C_{l \ominus s}) > \mathrm{wt}(\mathbf{e})$. Since we have the bound $d(C_l) \geq d_\varphi(l) = \min\{\nu_m | m \geq l, C_m \neq C_{m+1}\}$ where $\nu_l = \#\{(i,j) \in \mathbb{N}^2 | i \oplus j = l+1\}$, we want to correct t errors whenever $2t < d(C_l)$ which means that the inequality $d(C_{l \ominus s}) > \mathrm{wt}(\mathbf{e})$ is not always satisfied. It is exactly in this situation that the majority voting enters and we will shortly see how the algorithm described above can be extended to cover this case. We will first prove

Lemma 7.17 *If* $\mathrm{wt}(\mathbf{e}) = t$, *then for each* l, $\#\Delta_l \leq t$.

Proof. Let $s \in \Delta_l$. Then there exists an $f \in R$ such that $\rho(f) = \rho_s$. Let $[f]$ denote the class of f in R/L, where L is the ideal of error-locators as defined in the previous section. Then $[f] \neq [0]$ since $f \notin L$. Since the functions corresponding to the elements of Δ_l have different orders, they are linearly independent and the same is true for their classes in R/L. Therefore $\#\Delta_l \leq \dim(R/L) = t$, by the remark in Section 6. □

Let $\Gamma_l = \{s \in \Sigma_{l-1} | s \leq_p l, l \ominus s \in \Sigma_{l-1}\}$. If $s \in \Gamma_l$ then Lemma 7.4 implies that either s and $l \ominus s$ are elements of Σ_l or if a function of order ρ_s has fail l then s and $l \ominus s$ are both in Δ_l. Thus the increase of the delta set from $l-1$ to l is $\Gamma_l \cap \Delta_l$. The following proposition says that this increase is less than half of Γ_l if $2t$ is less than ν_l.

Proposition 7.18 *If* $\nu_l > 2t$, *then* $\#(\Gamma_l \cap \Sigma_l) > \#(\Gamma_l \cap \Delta_l)$.

Proof. Let $V_l = \Gamma_l \cap \Sigma_l$ and $W_l = \Gamma_l \cap \Delta_l$, these partition Γ_l. We note that W_l and Δ_{l-1} are disjoint subsets of Δ_{l-1}, and using Lemma 7.17, we therefore have

$$\#\Delta_{l-1} + \#W_l \leq t$$

On the other hand the four sets

$$V_l,\ W_l,\ \{a \in \Sigma_{l-1} | l \ominus a \in \Delta_{l-1}\} = A_l,\ \text{and}\ \{a \in \Delta_{l-1} | a \leq_p l\} = B_l$$

partition the set $\{a \leq_p l\}$. Now $\#\{a \leq_p l\} = \nu_l$ and the sets A_l and B_l contain at most $\#\Delta_{l-1}$ elements so we have

$$\nu_l \leq \#V_l + \#W_l + 2\#\Delta_{l-1}$$

which gives the result. □

Assume $2t < d_\varphi(l) \leq \nu_l$. We will use Proposition 7.1 together with the output of the algorithm described before to calculate $S(f_{l+1})$ from $S(f_i)$, $i \leq l$.

For each $a \in \Gamma_{l+1}$ choose an $s \in \sigma_l$ such that $s \leq_p a$. Let $f^{(s)} = F_l(s)$ and choose (the unique) $\omega \in \mathbb{F}_q^*$ such that $g = f_{l+1} + \omega f^{(s)} f_{(l+1)\ominus s}$ satisfies $\rho(g) < \rho_{l+1}$. We then have $S(f_{l+1}) = S(g) - S(\omega f^{(s)} f_{l+1-s})$. Let the vote by a for $S(f_{l+1})$ be $S(g)$. The proposition says that most of the time $S(\omega f^{(s)} f_{l+1-s}) = 0$ so that the majority of the $a \in \Gamma_{l+1}$ will vote for the correct answer. The same process may now be applied to $l+2$ and so on until we have sufficiently many syndromes.

Remark 7.19 We note that the two inequalities above correspond to the ineqalities $K + F \leq t$ and $\nu_l \leq T + F + 2K$ of the previous section.

Example 7.20 This is a continuation of Example 7.16. We will determine $S(f_{21})$.

$\Gamma_{21} = \{6, 8, 11\}$

6: $g = x^4y^2 + x^4(y^2 + \omega^{10}xy + \omega^{13}y + \omega^{13}x + \omega^{11})$ and $S(g) = \omega^6$
8: $g = x^4y^2 + x^2y(x^2y + \omega^4xy + \omega^3x^2 + \omega y + \omega^7x + \omega^4)$ and $S(g) = \omega^6$
11 : $g = x^4y^2 + y^2(x^4 + \omega^3x^3 + \omega^9x^2y + \omega^4x^2 + \omega^{14}xy + \omega^6y + \omega^5x + 1)$ and $S(g) = \omega^6$,

so $S(f_{21}) = \omega^6$.

Since the vote was unanimous the sets F and G remain unchanged.

In the same way we get $S(f_{22}) = \omega^3$ also unanimous. To determine $S(f_{23})$ we note that $\Gamma_{23} = \{6, 8, 9, 10, 13\}$.

6: $g = x^2y^4 + x^2y^2(y^2 + \omega^{10}xy + \omega^{13}y + \omega^{13}x + \omega^{11})$ and $S(g) = \omega^6$
8: $g = x^2y^4 + xy^3(x^2y + \omega^4xy + \omega^3x^2 + \omega y + \omega^7x + \omega^4)$ and $S(g) = \omega^6$
9: $g = x^2y^4 + xy^2[x(y^2 + \omega^{10}xy + \omega^{13}y + \omega^{13}x + \omega^{11})]$ and $S(g) = \omega^6$
10: $g = x^2y^4 + x^2y[y(y^2 + \omega^{10}xy + \omega^{13}y + \omega^{13}x + \omega^{11})]$ and $S(g) = \omega^6$
13: $g = x^2y^4 + y^2[x^2(y^2 + \omega^{10}xy + \omega^{13}y + \omega^{13}x + \omega^{11})]$ and $S(g) = \omega^6$,

so $S(f_{23}) = \omega^6$. It turns out that for $l = 24, 25, 26$ again there is only one value and we get $S(f_{24}) = \omega^4$, $S(f_{25}) = \omega^{11}$, $S(f_{26}) = \omega^4$ and the F- and G-sets are the same as before.

$\Gamma_{27} = \{6, 8, 9, 10, 11, 12, 13, 14, 17\}$. Here, however, there is one wrong vote corresponding to

11: $\quad g = x^3(x^5+y) + x^4(x^4+\omega^9x^2y+\omega^3x^3+\omega^{14}xy+\omega^4x^2+\omega^6y+\omega^5x+1)$

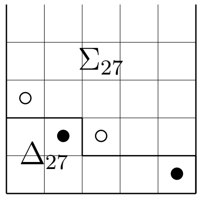
but others give $S(f_{27}) = \omega$. We finally get

$F : \{y^2+\omega^{10}xy+\omega^{13}y+\omega^{13}x+\omega^{11}, x^2y+\omega^4xy+\omega^3x^2+\omega y+\omega^7x+\omega^4\}$

and

$G : \{xy+\omega^{13}y+\omega^{14}x+\omega, x^4+\omega^9x^2y+\omega^3x^3+xy+\omega^4x^2+\omega^{11}y+\omega x+\omega\}$

and the corresponding figure (see Example 7.9)



So indeed $\#\Delta_{27} = 7$ and since

$$y^2+\omega^{10}xy+\omega^{13}y+\omega^{13}x+\omega^{11} = (y-\omega^3)(y+\omega^{10}x+\omega^8)$$

and

$$x^2y+\omega^4xy+\omega^3x^2+\omega y+\omega^7x+\omega^4 = (y-\omega^3)(x^2+\omega^4x+\omega)$$

it is easy to see that the common zeros are precisely the error points.

7.2 Computing the error values

In the special situation where we have a weight function ρ on an affine $\mathbb{F}_q$-algebra $R = \mathbb{F}_q[x_1,\ldots,x_m]/I$ with $\varphi : R \to \mathbb{F}_q^n$, the evaluation map $ev_{\mathcal{P}}$ where $\mathcal{P}$ consists of n distinct points of $\mathbb{F}_q^m$ in the zero set of I, and where the semigroup of nongaps is finitely generated, we will give a formula for determination of the error values after sufficiently many syndromes have been determined by the algorithm presented in Section 7.1.

Let $a_1, a_2, \ldots, a_\mu$ be a minimal set of generators and let Φ_j be the corresponding functions such that $\rho(\Phi_j) = a_j$. To any vector $\omega = (\omega_1, \ldots, \omega_\mu)$ of nonnegative integers corresponds the function

$$f_\omega = \prod_{s=1}^{\mu} \Phi_s^{\omega_s}$$

and

$$\rho(f\omega) = \sum_{s=1}^{\mu} a_s \omega_s.$$

Now

$$S_{\mathbf{e}}(f_\omega) = \sum_{j=1}^{n} e_j f_\omega(P_j) = \sum_{j=1}^{n} e_j \prod_{s=1}^{\mu} \Phi_s(P_j)^{\omega_s}$$

and suppose we know all syndromes $S_{\mathbf{e}}(f_\omega)$ where $0 \leq \omega_i \leq q-1$, $i = 1, \ldots, k$. Then for each P_e we can form the sum

$$\sum_{\omega} S_{\mathbf{e}}(f_\omega) \prod_{s=1}^{\mu} \Phi_s^{-\omega_s}(P_l)$$

where the summation is over all vectors ω with $1 \leq \omega_s \leq q-1$, $s = 1, \ldots, k$.

Inserting the expression for the syndromes, we get

$$\begin{aligned} \sum_{\omega} \sum_{j=1}^{n} e_j \prod_{s=1}^{\mu} \Phi_s^{\omega_s}(P_j) \Phi_s^{-\omega_s}(P_l) &= \sum_{j=1}^{n} e_j \prod_{s=1}^{\mu} \sum_{\omega} \left(\frac{\Phi_s(P_j)}{\Phi_s(P_l)} \right)^{\omega_s} \\ &= (-1)^k e_l \end{aligned}$$

and therefore e_l can be calculated. The last equality comes from the fact that if $\Phi_s(P_j) \neq \Phi_s(P_l)$, then

$$\sum_{\omega_s=1}^{q-1} \left(\frac{\Phi_s(P_j)}{\Phi_s(P_l)} \right)^{\omega_s} = 0.$$

If $j \neq l$, then for at least one s we have $\Phi_s(P_j) \neq \Phi_s(P_l)$ because otherwise the code would have minimum distance 2 and we will not consider such codes. Of course the calculation above is only valid if $\Phi_s(P_l) \neq 0$ for all $s = 1, \ldots, \mu$. If this is not the case, there is a slight modification which will give the error values.

The complexity in terms of the number of $\mathbb{F}_q$ additions and multiplications of the whole decoding procedure for the codes considered above is bounded above by

$$O(an^2) + O(q^{\mu+1}(a_1 + \cdots + a_\mu)) + O(n \cdot \mu \cdot q^\mu).$$

In the case of Hermitian curves this gives $O(n^{5/2})$.

7.3 Notes

The algorithm treated in this section is based on Sakata's extension [84, 85, 86] of the classical Berlekamp-Massey algorithm [9, 67]. The presentation of the algorithm in this section is an adaption of the paper [71]. The application of this algorithm to the decoding problem was made in [51] and the inclusion of the majority voting is from [87, 88, 89]. The algorithm was implemented in [65] for the Hermitian curve over $\mathbb{F}_{2^8}$ and a general implementation was carried out in [1]. In [55, 60] a generalization of Forneys formula is presented for the calculation of the error values. The formula in Section 7.2 is from [88]. Usually one can apply a fast Fourier-like transform to speed up this part of the decoding.

References

[1] J. Åberg, *An implementation of Sakata's algorithm,* Master's Thesis, Techn. Univ. Lund, Sweden, 1994.

[2] S.S. Abhyankar, *Algebraic geometry for scientists and engineers,* Mathematical Surveys and Monographs, vol. 35, Amer. Math. Soc., Providence 1990.

[3] S.S. Abhyankar, "Irreducibility criterion for germs of analytic functions of two complex variables," *Adv. Math.,* vol. 74, pp. 190-257, 1989.

[4] S.S. Abhyankar and T.T. Moh, "Newton-Puiseux expansion and generalized Tshirnhausen transformation," *J. Reine angew. Math.,* vol. 260, pp. 47-83, and vol. 261, pp. 29-54, 1973.

[5] S. Arimoto, "Encoding and decoding of p-ary group codes and the correction system," *Information Processing in Japan* (in Japanese), vol. 2, pp. 320-325, Nov. 1961.

[6] E.F. Assmus and ,J.D. Key, "Polynomial codes and finite geometries," *Handbook of Coding Theory,* V.S. PLess, W.C. Huffman and R.A. Brualdi, eds., Elsevier Amsterdam, 1998, pp. .....

[7] A. Barg, "Complexity issues in coding," *Handbook of Coding Theory,* V.S. PLess, W.C. Huffman and R.A. Brualdi, eds., Elsevier Amsterdam, 1998, pp. .....

[8] A.M. Barg, S.L. Katsman and M.A. Tsfasman, "Algebraic geometric codes from curves of small genus," *Probl. Inform. Trans.,* vol. 23, pp. 34-38, 1987.

[9] E.R. Berlekamp, *Algebraic coding theory*, McGraw-Hill, New York, 1968.

[10] R.E. Blahut, "Decoding of cyclic codes and codes on curves," *Handbook of Coding Theory,* V.S. PLess, W.C. Huffman and R.A. Brualdi, eds., Elsevier Amsterdam, 1998, pp. .....

[11] A.E. Brouwer, "Bounds on the size of linear codes," *Handbook of Coding Theory,* V.S. PLess, W.C. Huffman and R.A. Brualdi, eds., Elsevier Amsterdam, 1998, pp. .....

[12] B. Buchberger, *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal,* Ph.D. Thesis, Univ. of Innsbruck, Austria, 1965.

[13] C. Chevalley, *Introduction to the theory of algebraic function fields in one variable,* Amer. Math. Soc., New York 1951.

[14] D. Cox, J. Little and D. O'Shea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry*, Springer, Berlin 1992.

[15] P. Delsarte, J.-M. Goethals and F.J. MacWilliams, "On generalized Reed-Muller codes and their relatives," *Information and Control*, vol. 16, 403-442, 1970.

[16] V.G. Drinfeld and S.G. Vlăduţ, "Number of points of an algebraic curve," Func. Anal., vol. 17, pp. 53-54, 1993.

[17] I.M. Duursma, "Algebraic decoding using special divisors," *IEEE Trans. Inform. Theory*, vol. 39, pp. 694-698, March 1993.

[18] I.M. Duursma, ”Majority coset decoding,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 1067-1071, May 1993.

[19] I.M. Duursma, *Decoding codes from curves and cyclic codes,* Ph.D. Thesis, Eindhoven Univ. of Techn., Sept. 1993.

[20] D. Ehrhard, *Über das dekodieren algebraischer-geometrischer Codes,* Ph.D. Thesis, Univ. Düsseldorf, July 1991.

[21] D. Ehrhard, “Decoding algebraic-geometric codes by solving a key equation,” in AGCT-3, *Lect. Notes Math.,* vol. 1518, pp. 18-25, Springer, Berlin 1992.

[22] D. Ehrhard, ”Achieving the designed error capacity in decoding algebraic-geometric codes,” *IEEE Trans. Inform. Theory*, vol. 39, pp.743-751, May 1993.

[23] G.-L. Feng and T.R.N. Rao, “A class of algebraic geometric codes from curves in high-dimensional projective spaces,” in AAECC-10, *Lect. Notes in Comp. Sc.,* vol. 673, pp. 132-146, Springer, Berlin 1992.

[24] G.-L. Feng and T.R.N. Rao, ”Decoding of algebraic geometric codes up to the designed minimum distance,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 37-45, Jan. 1993.

[25] G.-L. Feng and T.R.N. Rao, ”A simple approach for construction of algebraic-geometric codes from affine plane curves,” *IEEE Trans. Inform. Theory*, vol. 40, pp. 1003-1012, July 1994.

[26] G.-L. Feng and T.R.N. Rao, ”Improved geometric Goppa codes,” Part I: Basic Theory, *IEEE Trans. Inform. Theory*, vol. 41, pp. 1678-1693, Nov. 1995.

[27] G.-L. Feng and T.R.N. Rao, ”Improved geometric Goppa codes,” Part II: generalized Klein curves, preprint 1994.

[28] G.-L. Feng and K. K. Tzeng, “A generalization of the Berlekamp-Massey algorithm for multisequence shiftregister synthesis with applications to decoding cyclic codes,” *IEEE Trans. Inform. Theory,* vol. 37, pp. 1274-1287, Sept. 1991.

[29] G.-L. Feng, V. Wei, T.R.N. Rao and K.K. Tzeng, "Simplified understanding and efficient decoding of a class of algebraic-geometric codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 981-1002, July 1994.

[30] G. Frey, M. Perret and H. Stichtenoth, "On the different of Abelian extensions of global fields," in AGCT-3, *Lect. Notes Math.*, vol. 1518, pp. 26-32, Springer, Berlin 1992.

[31] W. Fulton, *Algebraic curves. An introduction to algebraic geometry,* W.A. Benjamin Inc., New York Amsterdam, 1969.

[32] A. Garcia, S.J. Kim, and R.F. Lax, "Consecutive Weierstrass gaps and minimum distance of Goppa codes," *J. Pure Appl. Algebra*, vol. 84, pp. 199-207, 1993.

[33] A. Garcia and R. Lax, "Goppa codes and Weierstrass Gaps," in AGCT-3, *Lect. Notes Math.,* vol. 1518, pp. 33-42, Berlin 1992.

[34] A. Garcia and H. Stichtenoth, "Algebraic function fields over finite fields with many rational places," *IEEE Trans. Inform. Theory,* vol. 41, pp. 1548-1563, Nov. 1995.

[35] A. Garcia and H. Stichtenoth, "A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vlăduţ bound", *Invent. Math.*, vol. 121, pp. 211-222, 1995.

[36] A. Garcia and H. Stichtenoth, "On the asymptotic behaviour of some towers of function fields over finite fields," *Journ. Number Theory,* vol. 61, pp. 248-273, 1996.

[37] V.D. Goppa, "Codes associated with divisors," *Probl. Peredachi Inform.* vol. 13 (1), pp. 33-39, 1977. Translation: *Probl. Inform. Transmission*, vol. 13, pp. 22-26, 1977.

[38] V.D. Goppa, "Codes on algebraic curves," *Dokl. Akad. Nauk SSSR* vol. 259, pp. 1289-1290, 1981. Translation: *Soviet Math. Dokl.*, vol. 24, pp. 170-172, 1981.

[39] V.D. Goppa, "Algebraico-geometric codes," *Izv. Akad. Nauk SSSR*, vol. 46, 1982. Translation: *Math. USSR Izvestija*, vol. 21, pp. 75-91, 1983.

[40] V.D. Goppa, "Codes and information," *Russian Math. Surveys*, vol. 39, pp. 87-141, 1984.

[41] V.D. Goppa, *Geometry and codes*, Mathematics and its Applications, vol. 24, Kluwer Acad. Publ., Dordrecht, 1991.

[42] G. Haché, *Construction effective des codes géométriques,* Ph.D. Thesis, INRIA, Univ. Paris VI, Sept. 1996.

[43] J.P. Hansen, "Codes from the Klein quartic, ideals and decoding," *IEEE Trans. Inform. Theory*, vol. 33, pp. 923-925, Nov. 1987.

[44] R. Hartshorne, *Algebraic geometry,* Grad. Texts Math. vol. 52, Springer, Berlin 1972.

[45] C. Heegard, J. Little and K. Saints, "Systematic encoding via Gröbner bases for a class of algebraic-geometric Goppa codes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1752-1761, Nov. 1995.

[46] T. Henocq and D. Rotillon, "The theta divisor of a Jacobian variety and decoding of geometric Goppa codes," *J. Pure Appl. Algebra,* vol. 112, pp. 13-28, 1996.

[47] T. Høholdt and R. Pellikaan, "On the decoding of algebraic-geometric codes," *IEEE Trans. Inform. Theory,* vol. 41, pp. 1589-1614, Nov. 1995.

[48] Y. Ihara, "Some remarks on the number of rational points of algebraic curves of finite fields," *Journ. Fac. Sc. Univ. Tokyo* IA, vol. 28, pp. 721-724, 1981.

[49] T. Johnsen, S. Manshadi, and N. Monzavi, "A determination of the parameters of a large class of Goppa codes," *IEEE Trans. Inform. Theory,* vol. 40, pp. 1678-1681, Sept. 1994.

[50] J. Justesen, K.J. Larsen, H. Elbrønd Jensen, A. Havemose and T. Høholdt, "Construction and decoding of a class of algebraic geometric codes," *IEEE Trans. Inform. Theory*, vol. 35, pp. 811-821, July 1989.

[51] J. Justesen, K.J. Larsen, H. Elbrønd Jensen and T. Høholdt, "Fast decoding of codes from algebraic plane curves," *IEEE Trans. Inform. Theory*, vol. 38, pp. 111-119, Jan. 1992.

[52] T. Kasami, S. Lin and W.W. Peterson, "New generalization of Reed-Muller codes, Part 1: Primitive codes," *IEEE Trans. Inform. Theory,* vol. 14, pp. 189-199, 1968.

[53] C. Kirfel and R. Pellikaan, "The minimum distance of codes in an array coming from telescopic semigroups," *IEEE Trans. Inform. Theory*, vol. 41, no. 6, pp. 1720-1731, Nov. 1995.

[54] F. Klein, "Über die Transformation Siebenter Ordnung der Elliptischen Funktionen," *Math. Annalen,* vol. 14, pp. 428-471, 1879.

[55] R. Kötter, *On algebraic decoding of algebraic-geometric and cyclic codes,* Ph.D. Thesis, Linköping Univ., 1996.

[56] V. Yu. Krachkovskii, "Decoding of codes on algebraic curves," (in Russian), Conference Odessa, 1988.

[57] P.V. Kumar and K. Yang, "On the true minimum distance of Hermitian codes," in AGCT-3, *Lect. Notes Math.,* vol. 1518, pp. 99-107, Springer, Berlin 1992.

[58] G. Lachaud, "Les codes géométriques de Goppa," *Sem. Bourbaki*, no. 641, pp. 1-19, 1985.

[59] G. Lachaud, M.A. Tsfasman, J. Justesen and V.K.-W. Wei (Eds.), "Special issue on algebraic geometry codes," *IEEE Transactions on Information Theory,* vol. 41, Nov. 1995.

[60] D. Leonard, "A generalized Forney formula for AG codes," *IEEE Trans. Inform. Theory,* vol. 42, pp. 1263-1268, July 1996.

[61] J.H. van Lint, "Algebraic geometric codes," in Coding Theory and Design Theory, part I, *IMA Volumes Math. Appl.,* vol. 21, pp 137-162, Springer, Berlin, 1990.

[62] J.H. van Lint and G. van der Geer, *Introduction to coding theory and algebraic geometry*, DMV Seminar vol. 12, Birkhäuser Verlag, Basel Boston Berlin, 1988.

[63] J.H. van Lint and R.M. Wilson, "On the minimum distance of cyclic codes," *IEEE Trans. Inform. Theory*, vol. 32, pp. 23-40, Jan. 1986.

[64] B. López Jiménez, *Plane models of Drinfeld modular curves,* Ph.D. Thesis, Univ. Complutense, Madrid, March 1996.

[65] Y. Madelung, "Implementation of a Decoding Algorithm for AG-codes from the Hermitian Curve," Rep. IT-93-137, Technical University of Denmark, Lyngby, 1993.

[66] Yu.I. Manin and S.G. Vlăduţ, "Linear codes and modular curves," *Journ. Sov. Math.* vol. 30, pp. 2611-2643, 1985.

[67] J.L. Massey, "Shift-register synthesis and BCH decoding," *IEEE Trans. Inform. Theory*, vol. 15, pp. 122-127, Jan. 1969.

[68] S. Miura, "Algebraic geometric codes on certain plane curves," *IEICE Trans.,* vol. J. 75 A, no. 11, pp. 1739-1745, Nov. 1992.

[69] S. Miura and N. Kamiya, "Geometric-Goppa codes on some maximal curves and their minimum distance," in *Proc. 1993, IEEE Inform. Theory Workshop,* pp. 85-86.

[70] C. Moreno, *Algebraic curves over finite fields*, Cambridge Tracts in Math. vol. 97, Cambridge Univ. Press, Cambridge, 1991.

[71] M. E. O'Sullivan, "Decoding of Codes Defined by a Single Point on a Curve," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1709-1719, Nov. 1995.

[72] R. Pellikaan, "On a decoding algorithm for codes on maximal curves," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1228-1232, Nov. 1989.

[73] R. Pellikaan, "On the efficient decoding of algebraic-geometric codes," in Eurocode 92, *CISM Courses and Lectures,* vol. 339, pp. 231-253, Springer, Wien-New York 1993.

[74] R. Pellikaan, "The shift bound for cyclic, Reed-Muller and geometric Goppa codes," in *Proc. of AGCT-4,* pp. 155-174, W. de Gruyter, Berlin 1996.

[75] R. Pellikaan, "On the existence of order functions," to appear in *Journ. Statistical Planning and Inference.*

[76] R. Pellikaan, M. Perret and S. G. Vlăduţ (Eds.), *Arithmetic, Geometry and Coding Theory,* Proc. of AGCT-4, Luminy 1993, W. de Gruyter, Berlin 1996.

[77] R. Pellikaan, B.-Z. Shen, and G.J.M van Wee, "Which linear codes are algebraic-geometric?," *IEEE Trans. Inform. Theory*, vol. 37, pp. 583-602, May 1991.

[78] R. Pellikaan, H. Stichtenoth and F. Torres, "Weierstrass semigroups in an asymptotically good tower of function fields," to appear in *Finite Fields Appl.*

[79] W.W. Peterson, "Encoding and error-correction procedures for the Bose-Chauduri codes," *IRE Trans. Inform. Theory*, vol. 6, pp. 459-470, 1960.

[80] S.C. Porter, *Decoding codes arising from Goppa's construction on algebraic curves,* Ph. D. Thesis, Yale Univ., Dec. 1988.

[81] S.C. Porter, B.-Z. Shen and R. Pellikaan, "On decoding geometric Goppa codes using an extra place," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1663-1676, Nov. 1992.

[82] K. Saints and C. Heegard, "On hyperbolic cascaded Reed-Solomon codes," in Proc. AAECCC-10, *Lect. Notes Comp. Sc.*, vol. 673, pp. 291-303, Springer, Berlin 1993.

[83] K. Saints and C. Heegard, "Algebraic-geometric codes and multidimensional cyclic codes: A unified theory and algorithms for decoding using Gröbner bases," *IEEE Trans. Inform. Theory*, vol. 41, no. 6, pp. 1733-1751, Nov. 1995.

[84] S. Sakata, "Finding a minimal set of linear recurring relations capable of generating a given finite two-dimensional array," *Journal of Symbolic Computation*, vol. 5, pp. 321-337, 1988.

[85] S. Sakata, "Extension of the Berlekamp-Massey algorithm to N dimensions," *Information and Computation*, vol. 84, pp. 207-239, 1990.

[86] S. Sakata, "Decoding binary 2-D cyclic codes by the 2-D Berlekamp-Massey algorithm," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1200-1203, July 1991.

[87] S. Sakata, J. Justesen, Y. Madelung, H. Elbrønd Jensen and T. Høholdt, "Fast Decoding of Algebraic Geometric Codes up to the Designed Minimum Distance," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1672-1677, Nov. 1995.

[88] S. Sakata, H. Elbrønd Jensen and T. Høholdt, "Generalized Berlekamp-Massey decoding of algebraic geometric codes up to half the Feng-Rao bound," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1762-1768, Nov. 1995.

[89] S. Sakata, J. Justesen, Y. Madelung, H. Elbrønd Jensen and T. Høholdt, "A fast decoding method of AG codes from Miura-Kamiya curves C_{ab} up to Half the Feng-Rao bound," *Finite Fields and their Applications* vol. 11, pp. 83-101, 1995.

[90] I.R. Shafarevich, *Basic algebraic geometry,* 2nd ed., Springer, Berlin 1994,

[91] B.-Z. Shen, *Algebraic-geometric codes and their decoding algorithm,* Ph.D. Thesis, Eindhoven Univ. Techn., Sept. 1992.

[92] B.-Z. Shen and K.K. Tzeng, "Generation of matrices for determining minimum distance and decoding of algebraic-geometric codes," *IEEE Trans. Inform. Theory,* vol. 41, pp. 1703-1708, Nov. 1995.

[93] A.N. Skorobogatov and S.G. Vlăduţ, "On the decoding of algebraic-geometric codes," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1051-1060, Nov. 1990.

[94] A.B. Sørensen, "Projective Reed-Muller Codes," *IEEE Trans. Inform. Theory,* vol. 37, pp. 1567-1576, Nov. 1991.

[95] H. Stichtenoth, "Über die Automorphismengruppe eines algebraischen Funktionenkörpers von Primzahlcharakteristik. Teil II: Ein spezieller Typ von Funktionenkörpern," *Arch. Math.,* vol. 24, pp. 615-631, 1973.

[96] H. Stichtenoth, "A note on Hermitian codes over GF(q^2)," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1345-1348, Nov. 1988.

[97] H. Stichtenoth, *Algebraic function fields and codes*, Universitext, Springer, Berlin 1993.

[98] H. Stichtenoth and M.A. Tsfasman (Eds.), *Coding theory and algebraic geometry,* Proc. of AGCT-3, Luminy 1991, Lect. Notes Math., vol. 1518, Springer, Berlin 1992.

[99] H.J. Tiersma, "Codes coming from Hermitian curves," *IEEE Trans. Inform. Theory*, vol. 33, pp. 605-609, July 1987.

[100] M.A. Tsfasman and S.G. Vlăduţ, *Algebraic-geometric codes*, Mathematics and its Applications vol. 58, Kluwer Acad. Publ., Dordrecht, 1991.

[101] M.A. Tsfasman, S.G. Vlăduţ and T. Zink, "Modular curves, Shimura curves and Goppa codes, better than Varshamov-Gilbert bound," *Math. Nachrichten*, vol. 109, pp. 21-28, 1982.

[102] S.G. Vlăduţ, "On the decoding of algebraic-geometric codes over GF(q) for q$\geq$ 16," *IEEE Trans. Inform. Theory*, vol 36, pp. 1461-1463, Nov. 1990.

[103] C. Voβ and T. Høholdt, "An explicit construction of a sequence of codes attaining the Tsfasman-Vlăduţ-Zink bound", *IEEE Trans. Inform. Theory,* vol. 43, pp. 128-135, Jan. 1997 .

[104] B.L. van der Waerden, *Einführung in der algebraische Geometrie*, Springer, Berlin 1966.

[105] R.J. Walker, *Algebraic Curves*, New York, Dover, 1950.

[106] C. Xing, "On Automorphism Groups of Hermitian Codes," *IEEE Trans. Inform. Theory*, vol. 41, no. 6, pp. 1629-1635, Nov. 1995.