Algebraic geometry and coding theory

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
Rector Magnificus, prof.dr. R.A. van Santen, voor een
commisie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op maandag 24 september 2001 om 16.00 uur

door

Pieter Hendrik Turdus Beelen

geboren te Kampen
Dit proefschrift is goedgekeurd door de promotoren:

prof.dr. H.C.A van Tilborg
en
prof.dr. H. Stichtenoth

Copromotor:

dr. G.R. Pelikaan
Contents

1 Introduction ........................................... 1
   1.1 Algebraic geometry and coding theory .......... 1
   1.2 Curves with many points ...................... 4
   1.3 Function fields and Hurwitz’s theorem ......... 5
   1.4 Two asymptotically good towers of function fields 7
   1.5 Overview ........................................ 9

2 Curves and Newton polygons ......................... 11
   2.1 Introduction ................................... 11
   2.2 Irreducibility .................................. 13
   2.3 The delta-invariant and the Newton diagram ... 17
   2.4 The genus and the Newton polygon ............. 24
   2.5 The delta-invariant in a degenerate case ..... 27
   2.6 Classification of trinomial curves .......... 32
   2.7 Curves with many rational points ............ 38

3 Semigroups ............................................. 47
   3.1 Introduction ................................... 47
   3.2 Gaps and differential forms .................. 49
   3.3 Semigroups of plane curves ................... 53
   3.4 Semigroups of plane curves of type II ....... 61
   3.5 Application to the first Garcia-Stichtenoth tower 65
   3.6 Gap-lifting in the Garcia-Stichtenoth towers ... 79

4 Missing functions for plane curves of type II .... 83
   4.1 Introduction ................................... 83
   4.2 Plane curves of type II ....................... 85

5 Elliptic curves and large prime numbers .......... 91
   5.1 Prime-generating elliptic curves ............. 91
   5.2 A primality test for certain elliptic curves .. 95
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3</td>
<td>The Wagstaff conjecture</td>
<td>99</td>
</tr>
<tr>
<td>5.4</td>
<td>Jacobians of prime order</td>
<td>104</td>
</tr>
<tr>
<td>6</td>
<td>Elliptic curves and pseudorandom sequences</td>
<td>109</td>
</tr>
<tr>
<td>6.1</td>
<td>Some facts about elliptic curves</td>
<td>109</td>
</tr>
<tr>
<td>6.2</td>
<td>Pseudorandom sequences from elliptic curves</td>
<td>112</td>
</tr>
<tr>
<td>6.3</td>
<td>A family of sequences having small crosscorrelations</td>
<td>123</td>
</tr>
</tbody>
</table>

Bibliography                                                                 | 125  |
Index                                                                                | 131  |
Samenvatting                                                                        | 133  |
Curriculum Vitae                                                                   | 135  |
Chapter 1

Introduction

Summary In this chapter an introduction will be given to the subject of algebraic geometry codes (AG codes). We will start by giving the relation between algebraic geometry and coding theory. After that we will elaborate on this relation and introduce some theory. Finally we will state the research problems which will be investigated in further chapters.

1.1 Algebraic geometry and coding theory

In the 1980s, V. D. Goppa found a way to obtain codes from algebraic curves, which eventually led to codes which improve on the Gilbert-Varshamov bound. In this section we will briefly introduce some algebraic geometry needed in order to be able to describe his ideas. We will start with RS codes as an example.

Example 1.1.1 Let $F_q$ be the finite field with $q$ elements. Usually, an RS code defined over $F_q$ is described by viewing it as a cyclic code with generator polynomial of the form $g(X) = \prod_{i=1}^{d-1} (X - \alpha^i)$ where the parameter $d$ is in fact the designed minimum distance of the code. See for example page 93 of (van Lint 1999). However, another way of introducing this code is by considering it as an evaluation code. Define $L(m)$ to be the vector space of polynomials over $F_q$ of degree less than or equal to $m$. Define the map $Ev$ from $L(m)$ to $F_q^{q-1}$ by

$$Ev(f) = (f(\alpha), f(\alpha^2), \ldots, f(\alpha^{q-1})),$$

where $\alpha$ is some primitive root of unity of the field $F_q$. Then the code

$$C(m) = Ev(L(m))$$

is isomorphic to an RS code of designed minimum distance $q - k$ and dimension $k = m + 1$ if $m < q - 1$ (see page 94 of (van Lint 1999)). We can include 0 as a point at which we evaluate the elements of $L(m)$. Then we obtain an extended RS code.
There is a way to include the point \([1 : 0]\) as an evaluation-point as well. Then we find a doubly extended RS code.

We can view the previous example in a more algebraic geometric setting. Denote by \(\mathbb{P}^1(\mathbb{F}_q)\) the projective line defined over the field \(\mathbb{F}_q\). We can identify the field \(\mathbb{F}_q\) with an affine part of \(\mathbb{P}^1(\mathbb{F}_q)\) in a natural way if we use coordinates. Then an element \(\beta\) of the field corresponds to the point \([\beta : 1]\) of the projective line. We can interpret the polynomials in the vector space \(L(m)\) as functions on this projective line, which are not defined at the point \([1 : 0]\). The function corresponding to the polynomial \(X^l\) is said to have a pole of order \(l\) at infinity. In fact the vector space of functions corresponding to the vector space \(L(m)\) consists exactly of those functions having a pole of order at most \(m\) at the point \([1 : 0]\) and being regular elsewhere.

We can generalize this to arbitrary algebraic curves. Let \(C\) be an algebraic curve defined over \(\mathbb{F}_q\) and let \(P\) and \(Q_1, \ldots, Q_n\) be \(\mathbb{F}_q\)-rational points on it. Denote by \(L(mP)\) the vector space of functions \(f\) on \(C\) such that \(f\) is regular everywhere except at the point \(P\), where it has a pole of order at most \(m\). Define an evaluation map \(E_{v_{Q_1, \ldots, Q_n}}\) from \(L(mP)\) to \(\mathbb{F}_q^n\) by

\[ E_v(f) = (f(Q_1), \ldots, f(Q_n)) \]

and define

\[ C(m, P; Q_1, \ldots, Q_n) = E_{v_{Q_1, \ldots, Q_n}}(L(mP)). \]

Of course the question is what the minimum distance of this code will be. The above construction gives rise to one-point AG codes. The reason for this terminology is that instead of looking at the space of functions \(L(mP)\), one can look at more general spaces of functions allowing poles at other points as well. For example if \(R\) is another \(\mathbb{F}_q\)-rational point of the curve \(C\) the space \(L(P + 2R)\) consists of those functions \(f\) such that \(f\) has a pole of order at most \(1\) at \(P\) and a pole of order at most \(2\) at \(R\) (and is regular elsewhere). One can use these vector spaces of functions to obtain good codes as well. The terminology needed to describe these spaces of functions is that of a divisor. A divisor \(D\) is a formal sum of points of \(C\), say \(D = \sum_{i=1}^{s} a_i P_i\), where the \(P_i\) are points of \(C\) and \(a_i \in \mathbb{Z}\). The degree of the divisor \(D\) is defined to be the sum of the \(a_i\). The space of functions \(L(D)\) consists of those functions \(f\) such that \(f\) has a pole of order at most \(a_i\) at the point \(P_i\) for all \(i\) and is regular elsewhere.

The main tool needed to estimate the parameters of an AG code is the following theorem which is known as Riemann’s theorem (see page 21 and 28 of (Stichtenoth 1993)). One can also use the stronger theorem of Riemann-Roch, but this is not necessary for the estimation of the minimum distance of AG codes. Riemann’s theorem, as well as Riemann-Roch’s theorem uses the genus of the curve. It is a natural number depending only on the curve \(C\). We will come back to the determination of the genus in the next chapter.
1.1 Algebraic geometry and coding theory

**Theorem 1.1.2** Let $C$ be an algebraic curve defined over a perfect field $\mathbb{F}$. For any divisor $D$ we have

$$\dim(L(D)) \geq \deg(D) + 1 - g,$$

while equality holds if $\deg(D) \geq 2g - 1$.

Another fact that we will need is the that $\deg(D) < 0$ implies that $L(D) = \{0\}$, see for example page 18 of (Stichtenoth 1993). This gives an estimate for the minimum distance of the code $C(m, P, Q_1, \ldots, Q_n)$. We formulate the result only for one point AG codes, but the following proposition can easily be generalized to more general AG codes.

**Proposition 1.1.3** Let $C$ be an algebraic curve defined over $\mathbb{F}_q$ and suppose that $P, Q_1, \ldots, Q_n$ are distinct $\mathbb{F}_q$-rational points. The code $C(m, P, Q_1, \ldots, Q_n)$ has length $n$. Its dimension $k$ satisfies

$$k = \dim(L(mP)) - \dim(L(mP - Q_1 - \cdots - Q_n))$$

and hence the evaluation map $Ev_{Q_1, \ldots, Q_n}$ is injective as long as $m < n$. For the minimum distance $d$ of the code we have

$$d \geq n - m.$$

**Proof:** The statement about the length of the code $C(m, P, Q_1, \ldots, Q_n)$ is clear from the definition. The statement about the dimension $k$ follows from the fact that the kernel of the evaluation map consists exactly of the vector space $L(mP - Q_1 - \cdots - Q_n)$. If $m < n$, we see that this kernel is trivial. Now we prove the statement about the minimum distance. Suppose that $f \in L(mP)$ has $n - d$ zeroes among the points $Q_1, \ldots, Q_n$, say in $Q_{i_1}, \ldots, Q_{i_{n-d}}$. Then $f \in L(mP - Q_{i_1} - \cdots - Q_{i_{n-d}})$. Hence this vector space is not trivial which implies $\deg(mP - Q_{i_1} - \cdots - Q_{i_{n-d}}) \geq 0$. $\square$

**Corollary 1.1.4** Let the notation be the same as in the above proposition. Suppose that $m < n$. Then we have

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

**Proof:** By the above proposition we have

$$d \geq n - k + \dim(L(mP)) - \dim(L(mP - Q_1 - \cdots - Q_n)) - m.$$

However, since $m < n$ we have $\dim(L(mP - Q_1 - \cdots - Q_n)) = 0$ and by Riemann’s theorem $\dim(L(mP)) \geq m + 1 - g$. $\square$

Hence for AG codes we have a lower bound for the minimum distance. Note that by Singleton’s bound we have $d \leq n - k + 1$. 
1.2 Curves with many points

A question that remained unanswered in the previous section is: “How long can you make an AG code?” If the genus is too high relative to the length of the code, we do not obtain a good code. Hence we would like to know how many $F_q$-rational points a curve defined over $F_q$ can have. This explains the following definition. We denote by $C(F_q)$ the set of points of $C$ defined over $F_q$.

**Definition 1.2.1** Define

$$N_q(g) = \max_{\mathcal{C}} \{ \# \mathcal{C}(F_q) \},$$

where $\mathcal{C}$ varies over all curves defined over $F_q$ with genus $g$.

Further define

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

Several bounds are known for these numbers. From the point of application to coding theory, lower bounds are useful. Usually lower bounds arise by more or less explicit constructions of curves, while upper bounds can be derived theoretically. For an excellent table of such bounds for small fields of characteristic 2 and 3 see (van der Geer and van der Vlugt 2000). For the proof of most upper bounds the following theorem is essential. It is due to Hasse and Weil.

**Theorem 1.2.2 (Hasse-Weil)** Let $\mathcal{C}$ be an algebraic curve defined over $F_q$ of genus $g$. There exist algebraic integers $\alpha_1, \ldots, \alpha_g$ of length $\sqrt{q}$, such that for all positive integers $e$ we have

$$\#(\mathcal{C}(F_{q^e})) = q^e + 1 - \sum_{i=1}^g (\alpha_i^e + \overline{\alpha_i^e}).$$

For a proof of this theorem see for example p.169 (Stichtenoth 1993). We fix some notation in the following definition.

**Definition 1.2.3** The numbers $\alpha_i$ and $\overline{\alpha_i}$ in the above theorem are called the Frobenius eigenvalues of the curve. Further the quantity

$$\sum_{i=1}^g (\alpha_i + \overline{\alpha_i})$$

is called the trace of $\mathcal{C}$.

As a shorthand notation for $\#(\mathcal{C}(F_{q^e}))$ we will use $C_e$ if it is clear which curve we are talking about. The above theorem implies that if we know for some fixed curve $\mathcal{C}$ of genus $g$ the numbers $C_1, \ldots, C_g$, we can calculate all other $C_e$. This is especially useful if the field of definition $F_q$ is small. Another immediate consequence of this theorem is the following upper bound on $N_q(g)$.
1.3 Function fields and Hurwitz’s theorem

Corollary 1.2.4 We have

\[ N_q(g) \leq q + 1 + 2g\sqrt{q}. \]

It turns out that for curves defined over \( \mathbb{F}_q \) with “small” genus this bound is a good one. More specifically if \( g \leq (q - \sqrt{q})/2 \) the Hasse-Weil bound works well, while for larger genera it can be improved. For example the following upper bound by Ihara improves on the Hasse-Weil upper bound for the genera mentioned above. For a proof see (Ihara 1981).

Proposition 1.2.5 (Ihara) We have

\[ N_q(g) \leq q + 1 + \frac{\sqrt{(8g + 1)q^2 + 4(q^2 - q)q - q}}{2}. \]

It turns out that for large genera this bound can still be improved on. This leads to the Oesterlé upper bounds. As a reference see for example (Thomas 1997). In this way an upper bound for the quantity \( A(q) \) can be obtained. The following proposition states this upper bound. For a proof see (Vlăduţ and Drinfeld 1983).

Proposition 1.2.6 We have

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

This bound is known as the Drinfeld-Vlăduţ bound.

Again we stress that from the point of view of coding theory, we need lower bounds. We will see in the next section that if \( q \) is a square, the above upper bound for \( A(q) \) is sharp.

1.3 Function fields and Hurwitz’s theorem

Denote by \( C \) an algebraic curve defined over some perfect field \( F \). Denote by \( \mathcal{F}(C) \) the field of algebraic functions on \( C \). Assume for example that an affine part of the curve \( C \) is given by the prime ideal \( I \) of polynomials in the variables \( X_1, \ldots, X_n \). Then the function field of \( C \) is given by the field of fractions of the ring \( F[X_1, \ldots, X_n]/I \). We denote by \( x_i \) the residue class of \( X_i \) modulo the ideal \( I \). Hence an element of \( \mathcal{F}(C) \) can be seen as a rational function in the \( x_i \). By a theorem in algebraic geometry two curves have the same function field if and only if they are birationally equivalent. Moreover for any function field there exists exactly one non-singular curve (up to isomorphism) which has this particular function field (see for example Chapter I, Section 6 of (Hartshorne 1977)). This correspondence is in fact a (contravariant) equivalence of categories. Hence the study of non-singular algebraic curves is the same as that of function fields. When talking about a curve, even when the particular representation we use at that moment has singularities, we will always have the unique non-singular curve in mind that is birationally equivalent to it.
Introduction

One of the problems one often faces is that of the calculation of the genus. In the next chapter we will look into this problem for general plane curves and obtain several results using the theory of Newton polygons. Another approach which works quite well uses Hurwitz’s theorem. Here we assume that we have a separable algebraic map from the curve of which we want to know the genus to another curve of which we know the genus. If we have a curve defined over a perfect field, the genus does not change if we extend the field of definition to an algebraically closed field. Hence we can always assume for the calculation of the genus of a curve defined over a perfect field, that the field of definition is algebraically closed. For a proof of the following theorem see Chapter IV, Section 2 of (Hartshorne 1977). For the formulation of this theorem we need the ramification divisor. This is a divisor which incorporates information about the covering of one curve by another. We will explain later how to calculate this for the cases we are interested in.

**Theorem 1.3.1** Let $C_1$ and $C_2$ be algebraic curves defined over an algebraically closed field of characteristic $p > 0$ of genus $g_1$ and $g_2$ respectively. Suppose that there exists a finite separable map

$$\phi : C_2 \rightarrow C_1.$$ 

Then we have

$$2g_2 - 2 = [\mathcal{F}_2 : \mathcal{F}_1](2g_1 - 2) + \deg(R),$$

where $R$ denotes the ramification divisor.

Of course this theorem is not very useful if one cannot calculate the ramification divisor. In practice however, this is no problem. In the case of an Artin-Schreier extension it can be formulated explicitly as shown in the following proposition (taken from p. 115 of (Stichtenoth 1993)). It is most convenient to formulate this proposition in terms of function fields. Note that the map $\phi$ in the above theorem gives rise to an inclusion of function fields

$$\phi^* : \mathcal{F}(C_1) \hookrightarrow \mathcal{F}(C_2).$$

Hence we can assume without loss of generality that $\mathcal{F}(C_1)$ is a subfield of $\mathcal{F}(C_2)$. We now look at some special types of extensions. It will include the Artin-Schreier extensions. A polynomial $a(T)$ is called additive over $F$ if for all $u$ and $v$ in $F$ we have $a(u + v) = a(u) + a(v)$.

**Proposition 1.3.2** Let $\mathcal{F}(C_1)$ be an algebraic function field of characteristic $p > 0$ defined over $\overline{\mathbb{F}_q}$. Let $a(T)$ be an additive polynomial over $\mathbb{F}_q$ whose roots are defined over $\overline{\mathbb{F}_q}$. Let $u \in \mathcal{F}(C_1)$ be such that for every point $P$ of $C_1$ there is an element $z \in \mathcal{F}(C_1)$ (depending on $P$) such that

$$v_P(u - a(z)) \geq 0.$$
1.4 Two asymptotically good towers of function fields

or

\[ v_P(u - a(z)) = -m \]

with \( m > 0 \) and \( m \not\equiv 0 \pmod{p} \). Define \( m_P = -1 \) in the first case and \( m_P = m \) in the second case, then \( m_P \) is well-defined. Let

\[ \mathcal{F}(\mathcal{F}_2) = \mathcal{F}(\mathcal{F}_1)(y), \]

where \( y \) satisfies

\[ a(y) = u, \]

and \( u \neq a(w) \) for all \( w \in \mathcal{F}(\mathcal{C}_1) \). Denote by \( g_1 \) (respectively \( g_2 \)) the genus of \( \mathcal{C}_1 \) (respectively \( \mathcal{C}_2 \)).

If there exists a point \( P \) of \( \mathcal{C}_1 \) such that \( m_P > 0 \), the curve \( \mathcal{C}_2 \) is absolutely irreducible. Suppose that the curve \( \mathcal{C}_2 \) is absolutely irreducible. The place \( P \) is unramified if and only if \( m_P = -1 \). Further the place is totally ramified if and only if \( m_P > 0 \). The ramification divisor is given by

\[ \sum_P (\deg(a(T)) - 1)(m_P + 1)P, \]

where the sum is taken over all points of \( \mathcal{C}_1 \).

We have

\[ g_2 = \deg(a(T))g_1 + \frac{\deg(a(T)) - 1}{2} \left( -2 + \sum_P (m_P + 1) \deg(P) \right). \]

An example of an additive polynomial over \( \mathbb{F}_q \) is \( a(T) = T^q - T \) or \( a(T) = T^q + T \). The first example gives for \( q = p \) rise to Artin-Schreier extensions. The second example we will encounter in the next section.

1.4 Two asymptotically good towers of function fields

We now quote some results concerning the explicit construction of a tower of function fields giving rise to a lower bound on \( A(q) \). The first result in this direction (see (Tsfasman, Vlăduţ, and Zink 1982)) was that a certain tower of modular curves \( (\mathcal{C}_i)_i \), defined over \( \mathbb{F}_{p^2} \), satisfies

\[ \lim_{i \to \infty} \#\mathcal{C}_i(\mathbb{F}_{p^2})/g(\mathcal{C}_i) = p - 1. \]

Hence, this tower attains the Drinfeld-Vlăduţ bound. As a result the authors of (Tsfasman, Vlăduţ, and Zink 1982) mention the fact that for \( p > 7 \), the resulting codes improve on the Gilbert-Varshamov lower bound. One of the problems when applying this result is that the curves \( \mathcal{C}_i \) are not given explicitly. However later
explicit descriptions of towers with the same asymptotic behaviour were found (see (García and Stichtenoth 1995) and (García and Stichtenoth 1996)). We will state some results in the next two theorems.

**Theorem 1.4.1** Define the tower of function fields \((\mathcal{F}_i)\) defined inductively by

\[
\mathcal{F}_1 = \mathbb{F}_{q^1}(x_1)
\]

and

\[
\mathcal{F}_{i+1} = \mathcal{F}_i(z_{i+1}),
\]

where

\[
z_{i+1} = z_i^{q} + z_i^{q-1}
\]

and

\[
z_{i+1} = x_{i+1} x_i \text{ for all } i \geq 1.
\]

This tower attains the Drinfeld-Vlăduţ bound.

The reason for the introduction of the variable \(z_{i+1}\) is that it then becomes apparent that the tower of function fields consists of successive Artin-Schreier like extensions. Hence the genus can be calculated using a proposition similar to Proposition 1.3.2. For more details the reader is referred to (García and Stichtenoth 1995).

**Theorem 1.4.2** Define the tower of function fields \((\mathcal{G}_i)\) defined inductively by

\[
\mathcal{G}_1 = \mathbb{F}_{q^1}(y_1)
\]

and

\[
\mathcal{G}_{i+1} = \mathcal{G}_i(y_{i+1}),
\]

where

\[
y_{i+1}^{q} + y_{i+1} = \frac{y_i^q}{y_i^{q-1} + 1}
\]

This tower attains the Drinfeld-Vlăduţ bound.

Note that we can embed \(\mathcal{F}_i\) in \(\mathcal{G}_{i+1}\) by sending \((y_1, y_2, \ldots, y_i)\) to \((z_2, z_3, \ldots, z_{i+1})\), since

\[
z_{i+1}^q + z_{j+1} = x_j^{q-1} = \frac{x_j^q}{x_j^{q-1}} = \frac{z_j^q}{z_j + z_j} = \frac{z_j^q}{z_j^{q-1} + 1}.
\]

For more details see (García and Stichtenoth 1996).

As a corollary we state
Corollary 1.4.3 Suppose that $q$ is a square. We have

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

It remains an open question what $A(q)$ is if $q$ is not a square. For direct applications we would like to know especially the value of $A(2)$. Lower bounds are known, but these are probably not very sharp. We quote a result by Serre (see (Serre 1983)).

Theorem 1.4.4 There exists a positive constant $c$ which does not depend on $q$, such that

$$A(q) > c \log q.$$  

1.5 Overview

We will now state which problems we will investigate in the remaining chapters. In Chapter 2 we will try to improve on the lower bounds for the numbers $N_q(g)$. We will gain much insight in the substitute-and-reduce construction. Also we will give some new results in the theory of singularities. In Chapter 3 we will investigate semigroups of plane curves and the curves arising in the first García-Stichtenoth tower. In Chapter 4 we will construct explicitly codes arising from a certain type of plane curves having two points at infinity. Chapters 5 and 6 deal with two applications of elliptic curves. In Chapter 5 we describe two families of prime numbers, akin to the Mersenne prime numbers, and give a fast primality test for them. It turns out that these primes are connected with certain elliptic curves. In Chapter 6 we describe another application of elliptic curves in the area of pseudorandom sequences. We will obtain sequences with good balance and autocorrelation.
Chapter 2

Curves and Newton polygons

Summary In this chapter we will apply the theory of Newton polygons to construct curves with many rational points. More specifically, we apply this theory to the construction of curves with many rational points mentioned in (Justesen, Larsen, Jensen, Havemose, and Hoholdt 1989) and will, as a result, get a better understanding of this construction. Further, we will give some results on the calculation of delta-invariants. Some of these are known in the literature, but seem to have been forgotten by most people. Especially Baker’s theorem deserves to be remembered, but is hardly known. For this reason we will prove Baker’s theorem from scratch (in a more general setting than the original theorem). Also some irreducibility criteria for bivariate polynomials are mentioned. Apart from some remarks in Sections 2.6 and 2.7 as well as the tables in Section 2.7, the contents of this chapter has been published in “Designs, codes and cryptography”, see (Beelen and Pellikaan 2000). This article is joint work with G.R. Pellikaan.

2.1 Introduction

There are several methods known to obtain curves with many rational points. See for example (Doumen 1999), (van der Geer and van der Vlugt 1999) or (Auer 1999). For tables of the best-known curves see (van der Geer and van der Vlugt 2000). Another construction is mentioned in (Justesen, Larsen, Jensen, Havemose, and Hoholdt 1989). We call it the substitute-and-reduce construction. The following example shows how this construction works.

Example 2.1.1 Consider the Klein quartic with homogeneous equation

\[ X^3Y + Y^3Z + Z^3X = 0. \]

This curve has the points \((1 : 0 : 0)\) and \((0 : 1 : 0)\) at infinity over any field. The affine equation is \(X^3Y + Y^3 + X = 0\). The origin is a point of this curve. If
(x, y) \in \mathbb{F}_q^2$ is a point of this curve with non-zero coordinates, then $x^7 = 1$. So

$$0 = x^3y + y^3 + x = x^3y + x^7y^3 + x = ((x^2y) + (x^2y)^3 + 1).$$

Let $t = x^2y$. Then $t^3 + t + 1 = 0$. Since $t^3 + t + 1$ has three zeros in $\mathbb{F}_q$, we see that the Klein quartic has $3.7 = 21$ rational points over $\mathbb{F}_q$ with non-zero coordinates.

As mentioned before, one of the motivations for constructing algebraic curves over finite fields with many rational points relative to the genus comes from Goppa’s application to error-correcting codes. Using an algebraic curve over a finite field $\mathbb{F}_q$ of genus $g$ with $n$ rational points, one can make $\mathbb{F}_q$-linear error-correcting codes of length $n$, dimension $k$ and minimum distance $d$ such that $k + d \geq n + 1 - g$, see (Goppa 1981; Høholdt, van Lint, and Pellikaan 1998; Stichtenoth 1993; Tsfasman and Vladut 1991). The curves that arise from the above construction give epicyclic codes (see (Blahut 1998)) that are particularly well suited for efficient decoding algorithms. Furthermore some of the curves we will find are given by bivariate polynomials representing designs, see (Pellikaan 1998).

**Construction 2.1.2** In general the substitute-and-reduce construction mentioned in (Justesen, Larsen, Jensen, Høavmose, and Høholdt 1989) works as follows. Let $f(T)$ be a univariate polynomial with coefficients in $\mathbb{F}_q$. Let $P$ be a (cyclic) subgroup of $\mathbb{F}_q^*$ of order $v$. Denote by $r$ the number of zeros of $f(T)$ that are in $P$. Let $k$ and $l$ be non-negative integers. Consider the bivariate polynomial $F(X, Y)$ that is obtained from $X^i f(X^k Y)$ by reducing modulo $v$ all exponents $i$ and $j$ of the monomials $X^i Y^j$ that appear in $X^i f(X^k Y)$. Then $F(X, Y) = 0$ defines a plane curve over $\mathbb{F}_q$ that has at least $vr$ points in $\mathbb{P}^2$. As a matter of fact, it is very natural in this construction to assume that $\deg(f(T)) < v$. In this case the zeros of $f(T)$ need not be in $P$, since all powers of $Y$ that occur in $X^i f(X^k Y)$ are less than $v$. For a more detailed discussion about the number of rational points see Remark 2.6.10 and Proposition 2.6.11 directly following it.

The curve with affine equation $F(X, Y) = 0$ arising from the univariate polynomial $f(T)$ in the above construction may be reducible and may have singularities. For non-singular plane curves Plücker’s well-known formula gives us the genus in terms of the degree. This was generalized by Baker (Baker 1893) for singular plane curves in terms of the Newton polygon of the defining polynomial.

In Section 2.2 we introduce the Newton polytope of a polynomial and give conditions that guarantee that the polynomial is (absolutely) irreducible. The case of trinomial curves is completely dealt with. A criterion of irreducibility in terms of the delta-invariant is given. In Section 2.3 the delta-invariant of a singular point is bounded in terms of the Newton diagram of the defining polynomial. A non-degeneracy condition guarantees that the bound is tight. In Section 2.4 Baker’s theorem is proved for arbitrary characteristic. In Section 2.5 we improve this theorem for a degenerate case (which we call $p$-degenerate) and apply it to an infinite sequence of curves containing the Suzuki curve as a special case.
2.2 Irreducibility

In Section 2.6 it is noted that several choices for $k$ and $l$ in the substitute-and-reduce construction give, up to birational equivalence, the same curve. The case that $f(T)$ is a trinomial is investigated and a classification is made. In Section 2.7 examples are given of univariate polynomials $f(T)$ that give rise to curves with many rational points relative to the genus. Several mistakes in the literature on the calculation of the genus and the number of rational points are corrected. We reproduced many best-known results regarding the number of rational points.

2.2 Irreducibility

Let $F$ be a field. Let

$$F(X_1, \ldots, X_m) = \sum_{i \in I} \alpha_i X^i$$

be a polynomial in $m$ variables, where $I$ is a finite subset of $\mathbb{N}_0^m$ and $\alpha_i \in F^*$ for all $i \in I$, and $X^i = X_1^{i_1} \cdots X_m^{i_m}$ for $i = (i_1, \ldots, i_m) \in \mathbb{N}_0^m$. When $m = 2$ we will write $(X, Y)$ instead of $(X_1, X_2)$.

**Definition 2.2.1** Denote by $\Gamma(F)$ the convex hull of the points $i \in I$ in $\mathbb{R}^m_+$. The set $\Gamma(F)$ is called the *Newton polytope* of $F$, or the *Newton polygon* if $m = 2$. The boundary of $\Gamma(F)$ is denoted by $\partial \Gamma(F)$. Let $A, B$ be two subsets of $\mathbb{R}^m$. Define the (Minkowski-)sum of $A$ and $B$ by $A + B = \{a + b | a \in A, b \in B\}$.

**Lemma 2.2.2** Let $G, H \in F[X_1, \ldots, X_m]$ with $F = GH$. Then

$$\Gamma(F) = \Gamma(G) + \Gamma(H).$$

**Proof:** See (Sturmfels 1996, Lemma 2.2) and (Gao 1998, Lemma 2.1). \hfill \Box

**Definition 2.2.3** A point in $\mathbb{R}^m$ is called *integral* if all of its coordinates are integral. A polytope in $\mathbb{R}^m$ is called integral if all of its vertices are integral. An integral polytope $C$ is called *integrally decomposable* if there exist integral polytopes $A$ and $B$ such that $C = A + B$ where both $A$ and $B$ have at least two points.

We have the following Irreducibility Criterion.

**Theorem 2.2.4** Let $F$ be any field and $F \in F[X_1, \ldots, X_m]$ be a non-zero polynomial such that $F$ is not divisible by any of the $X_i$. If the Newton polytope of $F$ is integrally indecomposable, then $F$ is absolutely irreducible.

**Proof:** This is a consequence of Lemma 2.2.2. See (Gao 1998). \hfill \Box

**Corollary 2.2.5** Let $F$ be an algebraically closed field. Let $F(X, Y) = \alpha X^a - \beta Y^b$, where $\alpha, \beta \in F^*$. Then $F$ is reducible in $F[X, Y]$ if and only if $a$ and $b$ are not relatively prime.
Proof: (See for instance (Ostrowski 1976, Theorem IX).) If the characteristic $p$ of $F$ divides $a$ and $b$, then $F = G^p$ for some $G \in F[X, Y]$. If $a$ and $b$ have a divisor $e > 1$ that is relatively prime to $p$, then $a = ea_1$ and $b = eb_1$ for some positive integers $a_1$ and $b_1$. Further, we have $a = \alpha_1^e$ and $b = \beta_1^e$ for some $\alpha_1, \beta_1 \in F$. Denote by $\xi$ a primitive $e$-th root of unity in $F$. We have

$$F(X, Y) = \prod_{i=1}^{e}(\alpha_1^i X^{a_1} - \xi^i \beta_1^i Y^{b_1}).$$

Hence $F$ is reducible in both cases. The converse is a consequence of Theorem 2.2.4, since the line segment between $(a, 0)$ and $(0, b)$ is integrally indecomposable if and only if $a$ and $b$ are relatively prime. □

Proposition 2.2.6 Let $\Gamma$ be an integrally indecomposable polytope in $\mathbb{R}^m$ contained in a hyperplane $H$. Let $P$ be a point not in $H$. Let $S$ be a set of integral points in the convex hull of $P$ and $\Gamma$. Then the convex hull of $S$ and $\Gamma$ is integrally indecomposable.

Proof: See (Gao 1998, Theorem 3.12). □

Definition 2.2.7 Let $a$ and $b$ be positive integers. Define $wdeg$ to be the weighted degree with weights $a$ and $b$ on bivariate polynomials. It is determined by assigning to $X$ (respectively $Y$) the weight $b$ (respectively $a$).

Corollary 2.2.8 Let $a$ and $b$ be positive, relatively prime integers. Let $wdeg$ be the weighted degree with weights $a$ and $b$. Let

$$F(X, Y) = \alpha X^a + \beta Y^b + G(X, Y),$$

where $\alpha, \beta \in F^*$ and $G(X, Y) \in F[X, Y]$. If $wdeg(G) < ab$, then $F(X, Y)$ is absolutely irreducible.

Proof: This is a consequence of Proposition 2.2.6. Another proof is given in (Hoholdt, van Lint, and Pellikaan 1998, Corollary 3.18). □

Proposition 2.2.9 Let $\Gamma$ be an integral polytope in $\mathbb{R}^m$ contained in a hyperplane $H$. Let $Q_1, \ldots, Q_n$ be the vertices of $\Gamma$. Let $P$ be an integral point not in $H$. Suppose that the greatest common divisor of all $Q_{ij} - P_j$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$ is 1. Then the convex hull of $P$ and $\Gamma$ is integrally indecomposable.

Proof: See (Gao 1998, Theorem 3.4). □

Remark 2.2.10 If the univariate polynomial $f(T)$ in the substitute-and-reduce construction has three terms, then the corresponding bivariate polynomial also has three terms. Consider the trinomial
2.2 Irreducibility

\[ \alpha X^a + \beta X^b Y^c + \gamma Y^d, \]

then the Newton polygon is a triangle with \((a,0), (b,c)\) and \((0,d)\) as vertices. By Proposition 2.2.9, this triangle is integrally indecomposable if and only if \(a, b, c\) and \(d\) have no common divisor. For the irreducibility of bivariate polynomials consisting of three or four monomials also see (Ostrowski 1976).

**Proposition 2.2.11** Let \(F(X,Y) = \alpha X^a + \beta X^b Y^c + \gamma Y^d\), where \(a, b, c\) and \(d\) are non-negative integers, and \(\alpha, \beta, \gamma\) are non-zero elements of an algebraically closed field \(F\). Assume that \((a,0), (b,c)\) and \((0,d)\) are three distinct points. In that case \(F(X,Y)\) is reducible if and only if \(ac + bd = ad\) or the characteristic of \(F\) divides the exponents \(a, b, c\) and \(d\).

**Proof:** If the characteristic \(p\) of \(F\) divides \(a, b, c\) and \(d\), then \(F = G^p\) for some \(G \in F[X,Y]\). Notice that \(ac + bd = ad\) if and only if the points \((a,0), (b,c)\) and \((0,d)\) are collinear.

Suppose that \(ac + bd = ad\). If one of the exponents is zero, then \(F(X,Y)\) is univariate in \(X\) or \(Y\) of degree at least 2, since \((a,0), (b,c)\) and \((0,d)\) are three distinct points. So \(F(X,Y)\) is reducible, since \(F\) is algebraically closed. Hence we may assume that the exponents are positive. Let \(e = \text{gcd}(a,b)\). Then \(a = ea_1\) and \(d = ed_1\) for some positive integers \(a_1\) and \(d_1\) such that \(\text{gcd}(a_1, d_1) = 1\). If \(ac + bd = ad\), then \(a_1 e_1 + b_1 d_1 = e_1 a_1 d_1\). So \(b = b_1 a_1\) and \(c = c_1 d_1\) for some positive integers \(b_1\) and \(c_1\), since \(a_1\) and \(d_1\) are relatively prime. Hence \(b = b_1 + c_1 \geq 2\). Let \(f(T) = \alpha + \beta T^{c_1} + \gamma T^e\). It follows that \(F(X,Y) = X^a f(Y^{d_1}/X^{a_1})\). Now \(f(T)\) is reducible in \(F[T]\), since \(\deg f(T) = e \geq 2\). Hence \(F(X,Y)\) is reducible. Hence \(F\) is reducible in both cases.

Now we will prove the converse. Suppose that \(ac + bd \neq ad\). Assume that \(ac + bd > ad\). (The case \(ac + bd < ad\) can be treated similarly.) Let \(n = ac + bd - ad\). If \(F\) is reducible in \(F[X,Y]\), then

\[ F^n(X,Y,T) = \beta X^b Y^c + (\alpha X^a + \gamma Y^d)T^n \]

is reducible in \(F[X,Y,T]\). Let \(K = F(X,Y)\). Let \(g = -\beta X^b Y^c/(\alpha X^a + \gamma Y^d)\). The polynomial \(T^n - g\) is reducible in \(K[T]\), since \(\beta X^b Y^c\) and \(\alpha X^a + \gamma Y^d\) are relatively prime in \(F[X,Y]\) and \(F\) is reducible in \(F[X,Y]\). By Kummer’s theorem ([Lang 1993, VI Theorem 9.1]), we see that \(n \geq 2\) and that there exists an integer \(l \geq 2\) dividing \(n\) and an \(f \in K\) such that \(f^l = g\). Hence there are relatively prime polynomials \(U, V \in F[X,Y]\), such that \(f = U/V\). We see that

\[ \beta X^b Y^c V^{-l} = -(\alpha X^a + \gamma Y^d)U^l, \]

Let \(p^e\) be the largest power of \(p\) dividing both \(a\) and \(d\). Then all irreducible factors of \(\alpha X^a + \gamma Y^d\) have the same exponent \(p^e\), as we have seen in the proof of Corollary 2.2.5, and are relatively prime to \(X\) and \(Y\). Comparing the irreducible factors on both sides of the equation gives that \(e > 0\), since \(l \geq 2\), and \(l, b\) and \(c\) are divisible by \(p\).
Finally we give a criterion for irreducibility using the delta-invariant.

**Definition 2.2.12** Let $\mathcal{X}$ be a reduced projective plane curve of degree $m$ defined over an algebraically closed field $\mathbb{F}$. So the defining equation has no multiple factors. Let $P$ be a point of $\mathcal{X}$. Let $\mathcal{O}_P$ be the local ring of all rational functions on $\mathcal{X}$ that are regular at $P$. Let $\mathcal{O}_P$ be the normalization (or integral closure) of $\mathcal{O}_P$. The **delta-invariant** of $P$ is defined by

$$\delta_P = \dim_{\mathbb{F}} \mathcal{O}_P/\mathcal{O}_P$$

Note that for a regular point $P$ the ring $\mathcal{O}_P$ is integrally closed. Hence for any regular point $P$ we have $\delta_P = 0$.

**Definition 2.2.13** Let $\mathcal{Y}$ and $\mathcal{Z}$ be two distinct reduced plane curves. Let $P$ be a point of both curves. We may assume that $P$ is a point of the affine plane. Let $G = 0$ and $H = 0$ be the affine defining equations of $\mathcal{Y}$ and $\mathcal{Z}$, respectively. The **intersection multiplicity** $I_P(\mathcal{Y}, \mathcal{Z})$ is defined by

$$I_P(\mathcal{Y}, \mathcal{Z}) = \dim_{\mathbb{F}} \mathbb{F}[X,Y]_P/(G,H),$$

where $\mathbb{F}[X,Y]_P$ is the local ring of rational functions of the affine plane that are regular at $P$.

**Proposition 2.2.14** Let $\mathcal{Y}$ and $\mathcal{Z}$ be two distinct reduced plane curves and denote by $P$ a point of both curves. We have

$$\delta_P(\mathcal{Y} \cup \mathcal{Z}) = \delta_P(\mathcal{Y}) + \delta_P(\mathcal{Z}) + I_P(\mathcal{Y}, \mathcal{Z})$$

**Proof:** See (Hironaka 1957) or (Buchweitz and Greuel 1980, Lemma 1.2.2).

In the following proposition we denote by $\sum_P \delta_P$ the summation over all points of a given curve. This sum is finite, since the number of singular points is finite and $\delta_P = 0$ for a non-singular point.

**Proposition 2.2.15** Let $\mathcal{X}$ be a reduced projective plane curve of degree $m$ over an algebraically closed field $\mathbb{F}$. If

$$\sum_P \delta_P < m - 1,$$

then $\mathcal{X}$ is irreducible. If $\mathcal{X}$ has no linear factors and

$$\sum_P \delta_P < 2m - 4,$$

then $\mathcal{X}$ is irreducible.
2.3 The delta-invariant and the Newton diagram

**Proof:** Suppose that the curve is reducible. Then \( \mathcal{X} = \mathcal{Y} \cup \mathcal{Z} \), where \( \mathcal{Y} \) is a curve of degree \( l \) and \( \mathcal{Z} \) is a curve of degree \( m - l \), and \( 0 < l < m \). Bézout’s theorem (Hartshorne 1977, Corollary 7.8) implies that

\[
\sum_p I_P(\mathcal{Y}, \mathcal{Z}) = l(m - l),
\]

where the sum is taken over all the points in the intersection of \( \mathcal{Y} \) and \( \mathcal{Z} \). Hence

\[
\sum_p \delta_P(\mathcal{X}) \geq \sum_p I_P(\mathcal{Y}, \mathcal{Z}) = l(m - l),
\]

by Proposition 2.2.14. However, \( l(m - l) \geq m - 1 \), which gives a contradiction, when \( \sum_p \delta_P(\mathcal{X}) < m - 1 \). If the curve has no linear factor and \( \sum_p \delta_P < 2m - 4 \), then we get a contradiction as well, since \( l(m - l) \geq 2(m - 2) \). Therefore the curve is irreducible in both cases. \( \square \)

2.3 The delta-invariant and the Newton diagram

The following theorem is an extension of Plücker’s formula for singular plane curves.

**Proposition 2.3.1** Let \( \mathcal{X} \) be an absolutely irreducible plane curve of degree \( m \). Then the genus of the non-singular model of \( \mathcal{X} \) is

\[
g = \frac{(m - 1)(m - 2)}{2} - \sum_p \delta_P.
\]

**Proof:** See (Hartshorne 1977, Chapter IV, Exercise 1.8). \( \square \)

Let \( \mathcal{X} \) be a reduced plane curve with defining equation \( F = 0 \). Suppose that \( P \) is a point of the curve. After a coordinate change we may assume that \( P = (0,0) \). The multiplicity \( r_P \) of \( \mathcal{X} \) at \( P \) is equal to the minimal degree of the homogeneous parts of \( F \). So

\[
F(X,Y) = \sum_{r \geq r_P} F_r(X,Y),
\]

where \( F_r \) is homogeneous of degree \( r \) and \( F_{r_P} \neq 0 \).

The blow-up \( B_P \) of the affine plane with center \( P = (0,0) \) is the surface in \( \mathbb{A}^2 \times \mathbb{P}^1 \) with defining equation \( Xu = YV \), where \( (X,Y) \) are the coordinates of \( \mathbb{A}^2 \) and \( (U : V) \) are the homogeneous coordinates of \( \mathbb{P}^1 \) (see pp 28-30 of (Hartshorne 1977)). The projection of \( B_P \) to \( \mathbb{A}^2 \) is an isomorphism outside \( P \), and \( P \) is replaced by a projective line called the exceptional curve. The points of this projective line correspond to the directions at \( P \). The blow-up is covered by two affine pieces with coordinates \( X_1 = V/U \) and \( Y_1 = Y \) if \( U \neq 0 \), and \( X_2 = X \) and \( Y_2 = U/V \) if \( V \neq 0 \). The inverse image of \( \mathcal{X} \) under this projection is the union of the exceptional curve.
and the \textit{strict transform} \( \tilde{X} \) of \( X \). On the first affine piece we have \( X = X_1 Y_1 \) and \( Y = Y_1 \), so \( F(X, Y) = Y_1^r F_1(X_1, Y_1) \) and \( F_1(X_1, Y_1) = 0 \) is the equation of the strict transform. And similarly on the second affine piece we have \( X = X_2 Y_2 \) and \( Y = X_2 Y_2 \), so \( F(X, Y) = X_2^r F_2(X_2, Y_2) \) and \( F_2(X_2, Y_2) = 0 \) is the equation of the strict transform. We call the points on the exceptional curve given by \( (U : V) = (0 : 1) \) or \( (U : V) = (1 : 0) \) \textit{special}, since they correspond to the origins of the first and second affine piece, respectively. This process of blowing up at a point can be repeated several times at the points of the strict transforms, and in this way we get a tree of blow-ups. The nodes of the tree are the centers of the blow-ups and the intersection points of the strict transform with the exceptional curve. A center \( P \) is connected with a directed edge to a point \( Q \) on the strict transform directly above it, this is denoted by \( Q \rightarrow P \). We say that \( Q \) is an \textit{infinitely near point} of \( P \) if \( Q = P \) or there is a sequence \( Q_1, \ldots, Q_n \) such that \( Q = Q_1 \), \( P = Q_n \) and \( Q_i \rightarrow Q_{i+1} \) for all \( i < n \), this is denoted by \( Q \rightarrow P \). If the tree of blow-ups begins with \( P \) and ends with the points \( Q_1, \ldots, Q_m \) of the last strict transforms, then \( P \) is called the \textit{root} of the tree and \( Q_1, \ldots, Q_m \) the \textit{leaves} of the tree. If the multiplicities of the leaves are all one, then the curve is non-singular and the tree is called the \textit{desingularization} or \textit{resolution} tree.

\textbf{Remark 2.3.2} Let \( X \) be a reducible projective plane curve, then it has at least two components, and these components will intersect in a singular point of \( X \), as we have seen in the proof of Proposition 2.2.15, with at least two branches. Hence the desingularization tree of such a singular point has at least two leaves. This implies that if all singular points of a projective plane curve have exactly one branch, then the curve is absolutely irreducible. See (Abhyankar 1989).

\textbf{Proposition 2.3.3} Let \( X \) be a reduced plane curve defined over an algebraically closed field \( \mathbb{F} \). Let \( P \) be a point of \( X \), then there is a finite tree of blow-ups such that all leaves have multiplicity one and

\[ \delta_P = \sum_{Q \rightarrow P} \frac{r_Q(r_Q - 1)}{2}. \]

\textbf{Proof:} See (Hartshorne 1977, Chapter V, Example 3.9.3). \hfill \Box

\textbf{Corollary 2.3.4} If \( X \) is blown up once at \( P \) and the points on the exceptional curve are \( Q_1, \ldots, Q_n \), then

\[ \delta_P = \frac{r_P(r_P - 1)}{2} + \sum_{i=1}^{n} \delta_Q. \]

\textbf{Definition 2.3.5} Define

\[ \nu_P = \sum_{Q \text{ special}} \frac{r_Q(r_Q - 1)}{2}, \]

where the summation is taken over \( P \) and all special infinitely near points of \( P \).
2.3 The delta-invariant and the Newton diagram

**Remark 2.3.6** Clearly \( \delta_P \geq \nu_F \) by Proposition 2.3.3. Whereas \( \delta_P \) does not depend on the chosen coordinates, the number \( \nu_F \) does.

We will now introduce the Newton diagram of \( F \), relate it to \( \nu_F \) and show under which conditions it is equal to \( \delta_P \). Let

\[
F(X, Y) = \sum_{(m, n) \in \mathcal{I}} \alpha_{mn} X^m Y^n
\]

be a polynomial in two variables, where \( \mathcal{I} \) is a finite subset of \( \mathbb{N}_0^2 \) and \( \alpha_{mn} \in \mathbb{F}^* \) for all \( (m, n) \in \mathcal{I} \). Denote by \( \Gamma_+(F) \) the convex hull of the union for all \( (m, n) \in \mathcal{I} \) of the quadrants \( (m, n) + \mathbb{R}_+^2 \) in \( \mathbb{R}_+^2 \). Denote by \( \partial \Gamma_+(F) \), the union of the compact edges of \( \Gamma_+(F) \). The closure in \( \mathbb{R}_+^2 \) of the set \( \mathbb{R}_+^2 \setminus \Gamma_+(F) \) is denoted by \( \Gamma_-(F) \). Denote by \( \partial \Gamma_-(F) \), the boundary of \( \Gamma_+(F) \). The set \( \Gamma_-(F) \) is called the Newton diagram of \( F \). See (Baker 1893; Kouchnireenko 1975; Kouchnireenko 1976).

**Definition 2.3.7** Let \( F \) be a bivariate polynomial that is not divisible by either \( X \) or \( Y \). Write the main part \( F_{\text{main}} \) of \( F \) at \( P \) as

\[
F_{\text{main}}(X, Y) = \sum_{i=0}^{L} \beta_i X^{d_i} Y^{e_i}.
\]

That is to say that \( F_{\text{main}} \) consists of all terms \( X^{d_i} Y^{e_i} \) such that \( (d_i, e_i) \in \partial \Gamma_+(F) \) and \( \beta_i = \alpha_{d_i, e_i} \), and the \( d_i \) are increasing. Hence \( d_0 = 0 = e_L \), \( d_{i+1} > d_i \) and \( e_{i+1} < e_i \) for \( 0 \leq i < L \).

The boundary of \( \Gamma_-(F) \) consists of the horizontal edge \( \gamma_H \) from \((0,0)\) to \((d_L, 0)\), the vertical edge \( \gamma_V \) from \((0,0)\) to \((0,e_0)\), and edges \( \gamma_i \) between \((d_{i-1}, e_{i-1})\) and \((d_i, e_i)\), for \( i = 1, \ldots, L \), on the intersection of \( \Gamma_-(F) \) and \( \Gamma_+(F) \). See Figure 1.
Definition 2.3.8 Let \( \gamma \) be a line segment between the integral points \((a, b)\) and \((c, d)\). Define the arithmetic length \( l(\gamma) \) of \( \gamma \) by
\[
l(\gamma) = \gcd(a - c, b - d).
\]

Remark 2.3.9 The number of integral points on \( \gamma \) is equal to \( l(\gamma) + 1 \). Let \((a, b)\), \((c, d)\) and \((e, f)\) be three collinear integral points such that \( a < c < e \). Let \( \gamma_1 \) be the line segment between \((a, b)\) and \((c, d)\), \( \gamma_2 \) the line segment between \((c, d)\) and \((e, f)\), and \( \gamma \) the line segment between \((a, b)\) and \((e, f)\). Then \( l(\gamma) = l(\gamma_1) + l(\gamma_2) \).

Theorem 2.3.10 Let the notation be the same as in Definition 2.3.7. We have
\[
\nu_F = \text{area } \Gamma_-(F) + \frac{1}{2} \left( -l(\gamma_H) - l(\gamma_V) + \sum_{i=1}^{L} l(\gamma_i) \right).
\]

Proof: The proof is by induction on the number of special points. The number of special points is zero if and only if \( r_P = 1 \). If \( r_P = 1 \), then \( \nu_F = 0 \). Furthermore, \( L = 1 \) and \( \Gamma_-(F) \) is a triangle and \( d_1 = 1 \) or \( e_0 = 1 \). If \( d_1 = 1 \), then the area of \( \Gamma_-(F) \) is equal to \( e_0/2 \), \( l(\gamma_H) = 1 \), \( l(\gamma_V) = e_0 \) and \( l(\gamma_1) = 1 \). Hence the formula is valid. The case \( e_0 = 1 \) can be dealt with similarly.
2.3 The delta-invariant and the Newton diagram

Suppose $r_F > 1$. We will show the induction step. We will see what happens if we blow up $X$ at $P$. We know that $r_F$, the multiplicity of $P$, equals the degree of a monomial in $F$ of lowest degree. Hence there exist indices $j$ and $k$ such that $j \leq k$ and $d_i + e_i = r_F$ for all $j \leq i \leq k$. Take $j$ minimal and $k$ maximal with this property. The line with equation $m + n = r_F$ cuts the Newton diagram of $F$ into three parts. In Figure 1 these parts are denoted by $\Gamma_0$, $\Gamma_1$ and $\Gamma_2$.

After blowing up at $P$, the strict transform has (at most) two special points on the exceptional curve. Let $Q_1$ denote the first (possibly singular) point, with polynomial $F_1$, and $Q_2$ the second (possibly singular) point, with polynomial $F_2$. We have

$$\nu_F = \frac{r_F(r_F - 1)}{2} + \nu_{F_1} + \nu_{F_2}.$$ 

We can apply the induction hypothesis to $\nu_{F_1}$ and $\nu_{F_2}$, because $\nu_F > \nu_{F_1}$, $\nu_{F_2}$, since $r_F > 1$. Consider $Q_1$ with local coordinates $(X_1, Y_1)$, where $X = X_1 Y_1$ and $Y = Y_1$. If $F = \sum a_{mn}X^mY^n$, then $F_1 = \sum a_{mn}X^mY^{m+n-r_F}$. The map

$$(m, n) \mapsto (m, m + n - r_F)$$

is affine, since it is the composition of a special linear map in $SL_2(\mathbb{Z})$ and a translation. Hence this map preserves the area of domains and the arithmetic length of line segments.

Under this affine map the Newton diagram of $F_1$ corresponds to part $\Gamma_1$ of the Newton diagram of $F$ left to the point $(d_j, e_j)$. Similarly, the Newton diagram of $F_2$ corresponds to part $\Gamma_2$ of the Newton diagram of $F$ right to the point $(d_k, e_k)$. See Figure 1.

Using the induction hypothesis and the fact that the affine map preserves areas and arithmetic lengths we have

$$\nu_{F_1} = \text{area } \Gamma_1 + \frac{1}{2} \left( -l(\gamma_{H_1}) - (e_0 - r_F) + \sum_{i=1}^{j} l(\gamma_i) \right)$$

and

$$\nu_{F_2} = \text{area } \Gamma_2 + \frac{1}{2} \left( -(d_L - r_F) - l(\gamma_{V_2}) + \sum_{i=L+1}^{L} l(\gamma_i) \right),$$

where $\gamma_{H_1}$ is the horizontal edge of the Newton diagram of $F_1$ and $\gamma_{V_2}$ is the vertical edge of the Newton diagram of $F_2$. Now

$$l(\gamma_{H_1}) + \sum_{i=j+1}^{k} l(\gamma_i) + l(\gamma_{V_2}) = r_F,$$

since it is equal to the arithmetic length of the line segment between $(0, r_F)$ and $(r_F, 0)$ by Remark 2.3.9. Furthermore

$$\text{area } \Gamma \prec(F) = \text{area } \Gamma_0 + \text{area } \Gamma_1 + \text{area } \Gamma_2,$$
Curves and Newton polygons

while the area of $\Gamma_0$ is equal to $r^2_0/2$. Finally $l(\gamma_H) = d_L$ and $l(\gamma_V) = e_0$. Putting the above ingredients together gives the proof of the induction step. □

**Corollary 2.3.11** Let the notation be the same as in Definition 2.3.7. The number $\nu_F$ is equal to

$$
\sum_{i=0}^{N-1} (d_{i+1} - d_i) e_{i+1} + \frac{1}{2} \left((d_{i+1} - d_i - 1)(e_i - e_{i+1} - 1) + \gcd(d_{i+1} - d_i, e_i - e_{i+1}) - 1\right),
$$

which equals the number of unit-squares with integral vertices, that is so say sets of the form $(m, n) + [0, 1]^2$, $m, n \in \mathbb{N}_0$, contained in the set $\Gamma_-(F)$.

**Proof:** The first identity is a direct consequence of Theorem 2.3.10 after dividing the Newton diagram into rectangles with vertices $(d_i, 0)$, $(d_{i+1}, 0)$, $(d_i, e_i)$, and $(d_{i+1}, e_i)$, and the triangles with vertices $(d_i, e_i)$, $(d_{i+1}, e_i)$, and $(d_i, e_i)$. To prove the second identity, note that the terms in the summation are the numbers of unit squares in the rectangles and the triangles. The proof can be reduced to the case where $\Gamma'(F)$ consists of the triangle with vertices $(0, 0)$, $(a, 0)$ and $(0, b)$. The rectangle with vertices $(0, 0)$, $(a, 0)$, $(a, b)$ and $(0, b)$ consists of ab unit squares, and is the sum of twice the number of unit squares in the lower triangle plus those on the diagonal. See (Boden 1997). □

**Definition 2.3.12** Let $\gamma$ be a line segment of $\partial \Gamma_+(F)$ on the line with equation $ex + dy = de$. Hence there exist indices $j$ and $k$ such that $j < k$ and $cd_j + de_i = de$ for all $j \leq i \leq k$. Take $j$ minimal and $k$ maximal with this property. Define

$$F_\gamma(X, Y) = \sum_{i=j}^{k} \beta_i X^{d_i} Y^{e_i}.$$

**Definition 2.3.13** We call $F$ non-degenerate with respect to its Newton diagram, if $F$ is not divisible by either $X$ or $Y$, and for every line segment $\gamma$ of $\partial \Gamma_+(F)$ the ideal generated by $X \partial F \partial X$, $Y \partial F \partial Y$ and $F_\gamma$ has no zeros in $(\mathbb{F})^2$. The polynomial $F$ is called non-degenerate in the strong sense (Kouchnirevko 1975; Kouchnirevko 1976) with respect to its Newton diagram if the ideal $X \partial F \partial X$ and $Y \partial F \partial Y$ has no zeros in $(\mathbb{F})^2$ for every line segment $\gamma$ of $\partial \Gamma_+(F)$.

**Remark 2.3.14** Notice that both definitions of “non-degenerate” coincide in characteristic zero since we have Euler’s equation:

$$wF = aX \partial F \partial X + bY \partial F \partial Y$$
2.3 The delta-invariant and the Newton diagram

for a weighted homogeneous polynomial $F$ of weighted degree $w$, where $X$ has weight $a$ and $Y$ has weight $b$. In characteristic $p$ our definition is weaker. For instance the polynomial

$$X^{p+1} + X^{p-1}Y + XY^{p-1} + Y^{p+1}$$

is non-degenerate in our sense if the characteristic $p$ is not 2, but is degenerate in the strong sense.

**Remark 2.3.15** Let $d = d_k - d_j$ and $e = e_j - e_k$. Let $t = \gcd(d, e)$. Define $d' = d/t$ and $e' = e/t$. Then $F_\gamma$ is weighted homogeneous of weighted degree $de$, where $X$ has weight $e$ and $Y$ has weight $d$. Hence there exists a unique univariate polynomial $f_\gamma(T) \in F[T]$ of degree $t$ such that

$$F_\gamma(X, Y) = X^{d_k} Y^{e_k} f_\gamma(X^{-d'} Y^{-e'}).$$

It is easily verified that $F$ is non-degenerate with respect to its Newton diagram if and only if the polynomial $f_\gamma(T)$ has only simple zeros in $F$ for every line segment $\gamma$ of $\partial T_+(F)$. Interchanging the role of $X$ and $Y$ gives another way to define $f_\gamma$, but this gives the reciprocal polynomial of the previous $f_\gamma$.

Let $P$ denote the point $(0, 0)$ and $\delta_P$ the delta-invariant of the point $P$. The following proposition was anticipated in (Beden 1997).

**Proposition 2.3.16** Let $F$ be non-degenerate with respect to its Newton diagram. Then

$$\delta_P = \nu_F.$$ 

**Proof:** After blowing up the curve corresponding to $F$, at $P$, the strict transform has at most $d_k - d_j + 2$ points on the exceptional curve. Two of them are the special points $Q_1$ and $Q_2$. If $j < k$, the remaining points are regular points corresponding to the linear factors of $F_\gamma$, where $\gamma$ is the line segment on the line $m + n = r_P$. Note that by Remark 2.3.15 $F_\gamma$ splits into distinct linear factors, since $F$ is non-degenerate with respect to its Newton diagram. Following the proof of Theorem 2.3.10, we see that $F_1$ and $F_2$ are non-degenerate with respect to their Newton diagrams. Hence the only infinitely near points that may contribute to $\delta_P$ in the summation of the formula of Proposition 2.3.3 are $P$ and the special points. It follows that $\delta_P = \nu_F$. \qed

**Remark 2.3.17** Note that from the above proof we can also deduce that if $\gamma$ is a line segment of the Newton diagram with end-points $(a, b)$ and $(c, d)$, then the number of branches corresponding to this edge is exactly $\gcd(c - a, b - d)$. After a suitable number of blow-ups, this edge will become an edge making an angle of $\pi/4$ with both coordinate axes. The non-degeneracy now implies that after one more blow-up, the branches are separated from each other and that their number is as claimed. Whether these branches correspond to rational points depends on the coefficients of the defining polynomial $F$, i.e. the zeros in $F$ of the $f_\gamma$'s.
Remark 2.3.18 A formula for the genus of a plane curve over the complex numbers in terms of the Newton polygon of its defining equation can be found in (Baker 1983; Hovanskiĭ 1978). We will discuss this in the following section. The proof in (Baker 1983) implicitly uses the delta-invariant and is along the lines of this section. A generalization for complete intersection curves over the complex numbers is given in terms of the Newton polytopes of the defining equations in (Hovanskiĭ 1978).

Another proof of Theorem 2.3.10 could be obtained in characteristic zero with the help of the Milnor number $\mu_F$ of $F$, which is the codimension of the ideal generated by $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}$ in $\mathbb{F}[X, Y]$, the ring of formal power series in $X$ and $Y$. There exists a formula for the Milnor number in terms of the Newton diagram of $F$, when $F$ is non-degenerate, see (Kouchnirenko 1975; Kouchnirenko 1976). This holds in any characteristic. We get the desired result by applying Milnor’s equation (see (Buchweitz and Greuel 1980),(Milnor 1968))

$$\mu_F = 2\delta_F - \rho_F + 1,$$

where $\rho_F$ is the number of branches at $P$. This formula holds in characteristic zero but is not valid in characteristic $p > 0$. Take for instance

$$F_k = (1 + X^k)(X^p + Y^{p+1}),$$

where $k > 0$ and $k$ is not divisible by $p$, the characteristic of $\mathbb{F}$ (see (Greuel and Kröning 1990)). The partial derivatives are $kX^{k-1}(X^p + Y^{p+1})$ and $(1 + X^k)Y^p$. They generate the same ideal in $\mathbb{F}[X, Y]$ as $X^{p+k-1}$ and $Y^p$, since $1 + X^k$ is a unit in $\mathbb{F}[X, Y]$. Hence $\mu(F_k) = (p+k-1)p$. However, the (reducible) curve with equation $F_k = 0$ has the same delta-invariant $p(p-1)/2$ and only one branch at the origin for all $k$.

In (Saito 1983) a generalization of the delta-invariant of an isolated singularity of a hypersurface $F = 0$ of any dimension, and a formula for it in terms of the Newton diagram of $F$ is given.

2.4 The genus and the Newton polygon

Definition 2.4.1 We call $F(X, Y)$ non-degenerate with respect to its Newton polygon, if for every edge $\gamma$ of $\partial\Gamma(F)$ the ideal generated by $X\frac{\partial F}{\partial x}$, $Y\frac{\partial F}{\partial y}$ and $F_\gamma$ has no zeros in $\mathbb{F}$. The following theorem implies Baker’s theorem as a special case if we take $\mathbb{C}$ as the field of definition. For the case of the complex numbers, see (Baker 1983; Hovanskiĭ 1978). A similar theorem for the arithmetic genus in arbitrary characteristic was shown in (Kresch, Whetley, and Zieve 1999, Proposition 4.7).

Theorem 2.4.2 Let $F(X, Y) = 0$ define an irreducible curve $\mathcal{X}$ over an algebraically closed field. Then the genus $g$ of the non-singular model of $\mathcal{X}$ satisfies
2.4 The genus and the Newton polygon

\[ g \leq 1 + \text{area } \Gamma(F) - \frac{1}{2} \left( \text{number of integral points on } \partial \Gamma(F) \right). \]

This last expression is equal to the number of integral points in the interior of \( \Gamma(F) \).

Equality holds if \( F \) is non-degenerate with respect to its Newton polygon and the singular points of the homogeneous curve with equation \( F^*(X, Y, Z) = 0 \) are among \((0 : 0 : 1), (0 : 1 : 0) \) and \((1 : 0 : 0) \).

**Proof:** Suppose that \( F \) has degree \( d \). Consider the Newton polygon of \( F \) in Figure 2. It lies below the line with equation \( m + n = d \). Let \( P = (0 : 0 : 1), Q = (0 : 1 : 0) \) and \( R = (1 : 0 : 0) \). The triangle with vertices \((0, 0), (d, 0) \) and \((0, d) \) is divided into four parts: \( \Gamma_P, \Gamma_Q, \Gamma_R \) and \( \Gamma(F) \), where the first three parts are isomorphic to the Newton diagrams of \( F^*(X, Y, 1), F^*(X, 1, Z) \) and \( F^*(1, Y, Z) \) at \( P, Q \) and \( R \), respectively, under affine transformations that preserve areas and arithmetic lengths. Using Theorem 2.3.10, Proposition 2.3.1 and Remarks 2.3.6 and 2.3.9 we get the required inequality.

![Figure 2](image)

The statement on the number of integral points in the interior of the Newton polygon is shown similarly as Corollary 2.3.11. If \( F \) is non-degenerate with respect to its Newton polygon, then \( F^*(X, Y, 1), F^*(X, 1, Z) \) and \( F^*(1, Y, Z) \), are non-degenerate with respect to their Newton diagrams, so \( \delta_P = \nu_{F^*(X, Y, 1)}, \delta_Q = \nu_{F^*(X, 1, Z)} \) and \( \delta_R = \nu_{F^*(1, Y, Z)} \). If there are no singular points outside \( P, Q \) and \( R \) we get an equality in the upper bound for \( g \). \( \square \)
Corollary 2.4.3 Let $F$ be an algebraically closed field. Let $a$, $b$, $c$ and $d$ be non-negative integers. Suppose a curve is given by the equation

$$\alpha X^a + \beta X^b Y^c + \gamma Y^d = 0,$$

where $\alpha, \beta, \gamma \in F^*$, and $ac + bd \neq ad$, and $p$, the characteristic of $F$ does not divide each of the integers $a$, $b$, $c$ and $d$. Then the genus $g$ of the non-singular model of this curve satisfies

$$g \leq 1 + \frac{1}{2} \left(\left|ac + bd - ad\right| - \gcd(a - b, c) - \gcd(b, c - d) - \gcd(a, d)\right).$$

If $p$ does not divide $\gcd(a - b, c) \gcd(b, c - d) \gcd(a, d)(ac + bd - ad)$, then equality holds.

**Proof:** The curve is absolutely irreducible by Proposition 2.2.11, since $ac + bd \neq ad$ and $p$ does not divide each of $a$, $b$, $c$ and $d$. If $p$ does not divide $\gcd(a - b, c)$, $\gcd(b, c - d)$ and $\gcd(a, d)$, then the polynomial is non-degenerate with respect to its Newton polygon. Suppose that $S$ is a singular point not equal to one of the points $(0:0:1)$, $(0:1:0)$ or $(1:0:0)$. Then $S$ is not on the line at infinity. Hence $S = (x, y)$, where $xy \neq 0$ and $x$ and $y$ satisfy the equations $\alpha x^a + \beta x^b y^c + \gamma y^d = 0$, $\alpha x^{a-1} + \beta x^{b-1} y^c = 0$ and $c\beta x^b y^c + \gamma y^d = 0$. Hence we have

$$\begin{cases}
\alpha x^a + \beta x^b y^c + \gamma y^d &= 0, \\
\alpha x^a + \beta x^b y^c &= 0, \\
c\beta x^b y^c + \gamma y^d &= 0.
\end{cases}$$

This is a system of three linear equations in $\alpha x^a$, $\beta x^b y^c$ and $\gamma y^d$ with a non-trivial solution. The determinant of the corresponding $3 \times 3$ matrix is equal to $ac + bd - ad$, which is not zero in the field, since $ac + bd - ad$ is not divisible by the characteristic. This is a contradiction. Hence there is no singular point outside $(0:0:1)$, $(0:1:0)$ or $(1:0:0)$. Hence we can apply Theorem 2.4.2. \qed

**Example 2.4.4** Let $F$ be an algebraically closed field with characteristic not dividing $mn$. Consider the plane curve with affine equation

$$\alpha X^m + \beta Y^n + \gamma = 0,$$

where $\alpha, \beta, \gamma \in F^*$. Proposition 2.2.11 and Corollary 2.4.3 imply that this curve is absolutely irreducible and has genus

$$g = \frac{(m - 1)(n - 1) + 1 - \gcd(m, n)}{2},$$

if the characteristic does not divide $mn$. See also (Stichtenoth 1993, Example VI.3.3).
Example 2.4.5 The plane curve with equation
\[ X^m Y^n + Y^m + X^n = 0 \]
has been studied for several reasons by many authors (see for example the introduction in (Bennami and Carbonne 1997)). As a consequence of Proposition 2.2.11 and Corollary 2.4.3 this curve is absolutely irreducible and has genus
\[ g = 1 + \frac{(m^2 - mn + n^2) - 3 \gcd(m, n)}{2}, \]
if the characteristic does not divide \((m^2 - mn + n^2) \gcd(m, n)\).

2.5 The delta-invariant in a degenerate case

We know by Proposition 2.3.3 that
\[ \delta_p = \sum_{Q \rightarrow p} \frac{r_Q(r_Q - 1)}{2}, \]
where the sum is taken over all infinitely near points. Consider first the case that the Newton diagram of some polynomial is non-degenerate. We know how to calculate the delta-invariant in this case. Hence we know how to calculate the sum \(\sum_{Q \rightarrow p} r_Q(r_Q - 1)/2\). Since \(r_Q = 1\) for the leaves of the desingularization tree, we need not include them in the sum. If the Newton diagram is degenerate, then the desingularization tree changes. It becomes “longer”. However, the first part of the tree is the same as in the non-degenerate case. Also the \(r_Q\) are the same here. Hence this tree can be thought of as having a non-degenerate part, with other trees connected to its leaves. To calculate the delta-invariant of a degenerate singular point, it therefore suffices to use the usual formula for the first part of the tree and then calculate the sum of \(r_Q(r_Q - 1)\) for the rest of the points. Using this approach, we will obtain an explicit formula for the delta-invariant in the following case.

Definition 2.5.1 Let \(F(X, Y)\) be a bivariate polynomial defined over a field \(F\) of characteristic \(p > 0\). Assume \(F\) has the form
\[ F(X, Y) = X^K Y^L \left( \alpha X^{a+dp^i} Y^{b+ep^j} + \gamma X^{c} Y^{d} + \sum \varepsilon_{I,J} X^{I} Y^{J} \right), \]
where \(\alpha, \beta, \gamma \in F^*, \varepsilon_{I,J} \in F\), and \(a, b, c, d, e, i, j, I, J, K, L\) are integers such that \(a, b\) and \(e\) are positive, \(K, L\) are non-negative, \(0 < i \leq j\), while \(p\) does not divide any of the numbers \(a, b, e\) nor divide both \(c\) and \(d\). Further we assume \(\gcd(a, b) = 1\). Define
\[ \Delta(I, J) = aJ + bp^{i-j} I - abep^j \]
and
\[ \Delta = \Delta(c, d). \]

Assume \( \Delta > 0 \) and

\[ \Delta(I, J) \geq \frac{p^i}{p^i - 1} \]

for all \( I, J \) such that \( \varepsilon_{I, J} \neq 0 \), while the line segment between \( (K + aep^i, L) \) and \( (K, L + bep^i) \) is an edge of the Newton diagram of \( F \). Further assume that the Newton diagram has at most two edges, and if there is a second one it is the line segment from one of the end-points of the first to \( (K + c, L + d) \). Then the first edge is called \( p \)-degenerate.

The fact that the point \( (K + c, L + d) \) does not lie on the \( p \)-degenerate edge, but in fact lies above this edge, follows from the assumption \( \Delta > 0 \). The assumption that the Newton diagram has at most two edges, implies that \( K \) or \( L \) equals zero. The terms \( X^i Y^j \) should be thought of as irrelevant higher order terms. This is expressed by the inequality \( \Delta(I, J) \geq \frac{\Delta p^i}{(p^i - 1)} \). The condition that \( p \) does not divide both \( c \) and \( d \) ensures that the higher order terms are immaterial for the delta-invariant, as we will see later.

**Theorem 2.5.2** With the notation as above, let \( F(X, Y) \) have a \( p \)-degenerate edge in its Newton diagram. Let \( p^i \) be the largest power of \( p \) that divides \( \Delta \). Then the degeneracy gives rise to the following extra contribution \( \delta_{\text{deg}} \) to the delta-invariant:

\[ \delta_{\text{deg}} = \frac{c}{2} \left( \Delta p^i - \Delta p^{-1} - p^i + 1 \right). \]

Before we can prove this theorem, we need to derive several lemmas. The first one deals with the following problem. Consider a bivariate polynomial and an edge of its Newton polygon. We know that after blowing up the corresponding singularity several times, the points corresponding to this edge lie on an exceptional curve. We say that we have isolated the edge. If the polynomial under consideration is non-degenerate with respect to its Newton polygon, all these points are regular. However, if it is not, these points might be singular and are hence of interest. Can the equation obtained from the original polynomial by several blow-ups be produced in one step? In other words: how can we find the rational map which is the composition of the blow-ups mentioned above? Note that in the last step of this chain of blow-ups, there are two possible choices, we could replace \( X \) by \( XY \) and keep \( Y \) fixed or we could replace \( Y \) by \( XY \) and keep \( X \) fixed. These two possibilities are given in the next lemma.

**Lemma 2.5.3** Let \( G(X, Y) \) be a bivariate polynomial and suppose that an edge \( \gamma \) of its Newton polygon joins the points \( (K, L) + (r', 0) \) and \( (K, L) + (0, s') \). Further assume that \( \gcd(r', s') = g' \) and define \( r = r'/g' \) and \( s = s'/g' \). Let \( R \) and \( S \) be the smallest non-negative integers such that \( rs - sR = 1 \). The birational maps \( \phi^{-1} \) and \( \psi^{-1} \) given by
2.5 The delta-invariant in a degenerate case

\[ \phi(X, Y) = (X^r Y^{-s}, X^{-R} Y^S) \]

and

\[ \psi(X, Y) = (X^{-R} Y^{-s} + S, X^{-r} Y^s), \]

respectively, give equivalences between the curve defined by \( G(X, Y) = 0 \) and the one obtained from the first one by isolating \( \gamma \).

**Proof:** The mapping \( \phi \) should have as first coordinate \( X^r Y^{-s} \) or as second coordinate \( X^{-r} Y^s \), since \( \gamma \) has to be isolated. Say the first coordinate is \( X^r Y^{-s} \). Denote by \( \psi \) the map which has as second coordinate \( X^{-r} Y^s \). The other coordinate of \( \phi \) should have the form \( X^{-Rt + tr} Y^{S + ts} \) with \( R - S = t \), while \( t \) is some integer. We can assume that \( R \) and \( S \) are non-negative and minimal. Assuming this, we see that \( t \geq 0 \). From the particular choice of \( R \) and \( S \), we conclude that \( r - R \geq 0 \) and \( s - S \geq 0 \). Also we see that \( r - R \) and \( s - S \) are non-negative minimal solutions for the equation \( sR' - rS' = 1 \). Hence the map \( \psi \) (which has as second coordinate \( X^{-r} Y^s \)), has first coordinate of the form \( X^{r - R + ur} Y^{-(s - S + us)} \), where \( u \) is a non-negative integer. However, it follows by going back one step in the desingularization tree, that the map with first coordinate \( X^{r - R + ur} Y^{-(s - S + us)} \) and second coordinate \( X^{-(R + tr)} Y^{S + ts} \) is the map obtained by the composition of all blow-ups but the last one. Hence \( 1 = (r - R + ur)(S + ts) - (-(s - S + us))(- (R + tr)) = t + u \). This implies \( t + u = 0 \) and hence \( t = u = 0 \), since \( t \) and \( u \) are non-negative. \( \square \)

Now we apply Lemma 2.5.3 to our original situation.

**Lemma 2.5.4** Let the notation be the same as in Lemma 2.5.3. Using the map \( \phi \) or \( \psi \) mentioned above, we can isolate the edge \( \gamma \) connecting \( (K + aep^t, L) \) and \( (K, L + bep^t) \). If \( K = 0 \) we find, using \( \phi \), the defining polynomial

\[ \alpha X^{ep^t} + \beta + \gamma X^{S+Rd-Rbep^t} Y^\Delta + \sum \varepsilon_{I,J} X^{IS+RI-Rbep^t} Y^\Delta(I,J). \]

If \( L = 0 \) we find, using \( \psi \), the polynomial

\[ \alpha + \beta Y^{ep^t} + \gamma Y^{(b-S)+(a-R)d-(a-R)bep^t} X^\Delta + \sum \varepsilon_{I,J} Y^{I(b-S)+(a-R)J-(a-R)bep^t} X^\Delta(I,J), \]

**Proof:** The proof is straightforward and left to the reader. Use Lemma 2.5.3. \( \square \)

**Lemma 2.5.5** Suppose \( F(X, Y) \) is as above. Let \( \zeta_1, \ldots, \zeta_c \) be the elements in \( \overline{T} \) such that \( \alpha T^{ep^t} + \beta = \alpha \prod_{i=1}^c (T - \zeta_i)^{p^t} \). The equation obtained by using \( \phi \) has the form

\[ \alpha X^t^{p^t} + \mu Y^\Delta + \nu X^t Y^\Delta + \sum \omega_{I,J} X^{I^t} Y^{J} = 0, \]
in $X_i = X - \zeta_i$ and $Y$, while by using $\psi$ we find
\[ \beta Y_t^{p^j} + \mu' X^\Delta + \nu' Y_1 X^\Delta + \sum \omega'_{iJ} Y_t^I X^J = 0, \]
in $Y_i = Y - \zeta_i^{-1}$ and $X$. Here $\mu \mu' \neq 0$ and moreover, if $p|\Delta$, then $\nu \nu' \neq 0$. Furthermore $\Delta I + p^i J - p^i \Delta \geq \Delta p^i/(p^i - 1)$ for all $I$ and $J$ such that $\omega_{iJ} \neq 0$ or $\omega'_{iJ} \neq 0$.

Proof: By using Lemma 2.5.4 and taking proper care of the higher order terms, we can see that all statements are clear except the one about $\nu$ and $\nu'$. After some calculations, we see that $\nu = 0$ if and only if $p|cB + Ad - Abep^i$ and $\nu' = 0$ if and only if $p|c(b - B) + (a - A)d - (a - A)be p^i$. Let us consider the first case. We can assume $j > 0$, since otherwise $F(X, Y)$ is not degenerate and we are done (that is: we know the delta-invariant). Hence $\nu = 0$ if and only if $p|cB + Ad$. Recall that $\Delta = bcep^i - ad - abep^i$. Hence $p|\Delta$ implies $p|bcep^i - ad$. Also recall that $ab - bpe p^i A = 1$. Hence if $p|cB + Ad$ and $p|bcep^i - ad$, then $p|c$ and $p|d$. This gives a contradiction.

The statement about $\nu'$ can be proved analogously.

We are now ready to prove Theorem 2.5.2.

Proof: Write $l = ni + k$ where $k < i$ or $k = i = 0$. After applying the previous lemmas, we see that the singular point splits into $e$ points after using the “usual” sequence of blow-ups separating the edge $\gamma$. The Newton diagram of any of these points has one edge, connecting $(0, \Delta)$ and $(p^i, 0)$. Hence we find an extra contribution to the delta-invariant of $e((\Delta - 1)(p^i - 1) + p^i - 1)/2$. Another application of Lemma 2.5.3 (check that all conditions are fulfilled) gives a Newton diagram with one edge connecting $(0, \Delta/p^i)$ and $(p^i, 0)$. Hence we find an extra contribution to the delta-invariant of $e((\Delta/p^i - 1)(p^i - 1) + p^i - 1)/2$. This will continue until we find a Newton diagram with one edge connecting $(0, \Delta/p^i)$ and $(p^i, 0)$. This gives a contribution to the delta-invariant of $e((\Delta/p^i - 1)(p^i - 1) + p^i - 1)/2$. After one more application of the lemma we find an equation with a Newton diagram consisting of one edge connecting $(0, \Delta/p^i)$ and $(p^i, 0)$. Hence we have arrived at an equation which is non-degenerate with respect to its Newton diagram. Hence in total we find as extra contribution to the delta-invariant:
\[ \frac{e}{2} \left( (\Delta p^{-i} - 1)(p^i - 1) + p^k - p^i + \sum_{i=0}^{n} (p^i - 1)\Delta p^{-it} \right), \]
which is equal to
\[ \frac{e}{2} (\Delta p^i - \Delta p^{-i} - p^i + 1). \]

This concludes the proof.

Remark 2.5.6 For Theorem 2.5.2 to hold, the inequality $\Delta(I, J) \geq \Delta p^i/(p^i - 1)$ in the definition of $p$-degenerate can be weakened to the strict inequality.
2.5 The delta-invariant in a degenerate case

\[ \Delta(I,J) > \frac{p^i \cdot p^{-(n+1)i}}{p^i - 1} \Delta, \]

where \( n \) is the largest integer such that \( p^n \) divides \( \Delta \).

**Remark 2.5.7** Note that from the above we can conclude that corresponding to the edge \( \gamma \) of \( F(X,Y) \) (which connects \( (K + aep^i, L) \) and \( (K, L + bep^i) \)), there are \( e \) points in the non-singular model. These points are rational over the splitting field of \( T^e + (\beta/\alpha)^1/p^e \).

**Proposition 2.5.8** Consider the curve \( \mathcal{X}(q_0, q) \) with equation

\[ X^q - Y^{q_0 + q} + X^{q-1}Y^{q_0 + 1} - X^{q_0 + 1} = 0, \]

where \( q_0 \) and \( q \) are powers of \( p \), the characteristic of the field of constants. Then \( \mathcal{X}(q_0, q) \) is absolutely irreducible if and only if \( q_0 \neq \sqrt[q]{q} \). If \( q_0 \neq \sqrt[q]{q} \), then the genus \( g \) of the non-singular model of \( \mathcal{X}(q_0, q) \) is

\[ g = \begin{cases} \frac{q^2}{2} (q-1) & \text{if } q_0 < \sqrt[q]{q} , \\ \frac{q^2}{2} (q-1) & \text{if } q_0 > \sqrt[q]{q} . \end{cases} \]

**Proof:** Suppose first \( q_0 = \sqrt[q]{q} \). Then reducing the defining polynomial modulo \( Y^{q_0 + 1} - X - X^{q_0} \) gives zero. So the curve is reducible in this case.

Now suppose that \( q_0 \neq \sqrt[q]{q} \). Let \( P \) be the origin. Then \( P \) is the only singular point, we will see that its resolution tree has exactly one leaf. Hence the curve is absolutely irreducible by Remark 2.3.2.

The area of the Newton polygon \( \Gamma \) is \( (q^2 + q_0q - q_0 - 1)/2 \). The number of integral points on the boundary of \( \Gamma \) is \( 2q + \min\{q_0, q\} \) and the number of integral points in the interior of \( \Gamma \) is \( (q^2 + q_0q - 2q - q_0 - \min\{q_0, q\} + 2)/2 \). However, the edge between \( (q_0, 0) \) and \( (0, q_0 + q) \) of \( \Gamma \) is degenerate.

If \( q_0 < q \), then \( p^i = q_0 \) and \( \Delta(q_0 + q - 1, 0) = (q + q_0)(q - 1)/q_0 \) and \( \Delta(q - 1, q_0 + 1) = q(q - 1)/q_0 \). The edge between \( (q_0, 0) \) and \( (0, q_0 + q) \) is \( p \)-degenerate if and only if \( \sqrt[q]{q} < q_0 \).

If \( \sqrt[q]{q} < q_0 < q \), then \( \delta_{dg} = (q_0^2 - 2q - q_0 + 2)/2 \) and the genus of the non-singular model is \( g = q_0(q - 1)/2 \) by Theorem 2.5.2.

If \( q \leq q_0 \), then the degenerate edge is \( p \)-degenerate, \( \delta_{dg} = (q_0^2 - 3q_0 + 2)/2 \) and the genus of the non-singular model is \( g = q_0(q - 1)/2 \) by Theorem 2.5.2.

If \( q_0 < \sqrt[q]{q} \), then the degenerate edge is not \( p \)-degenerate. After the transformations \( (X,Y) \mapsto (XY^{1+q/q_0}, Y) \) and \( X \mapsto X - 1 \) we get a polynomial with degenerate edge between \( (q_0, 0) \) and \( (0, q(q - 1)/q_0) \). After the transformations \( (X,Y) \mapsto (XY^{q(q-1)/q_0}, Y) \) and \( X \mapsto X - 1 \) we get a polynomial with non-degenerate edge between \( (q_0, 0) \) and \( (0, q - 1) \). Hence

\[ \delta_p = \frac{(q_0 - 1)(q + q_0)}{2} + \frac{(q_0 - 1)(q - 1)}{2q_0} + \frac{(q_0 - 1)(q - 2)}{2} \]
and 
\[ g = \frac{1}{2}(q + q_0 - 1)(q + q_0 - 2) - \delta_P = \frac{q}{2q_0}(q - 1). \]

\[ \square \]

**Remark 2.5.9** Notice that the genera of the non-singular models of \( \mathcal{X}(q_0, q) \) and \( \mathcal{X}(q/q_0, q) \) are the same if \( q_0 < \sqrt{q} \). Another affine equation of the curve \( \mathcal{X}(q_0, q) \) is \( Z^q - Z = Y^{q_0}(Y^q - Y) \). Let \( q_0 < \sqrt{q} \). Let \( V = Y^{q/q_0+1} - Z^{q/q_0} \), then \( V^q - V = Y^{q/q_0}(Y^q - Y) \). Hence \( \mathcal{X}(q_0, q) \) and \( \mathcal{X}(q/q_0, q) \) are birationally equivalent and their non-singular models are isomorphic. This is a generalization of the proof of (Hansen and Stichtenoth 1990, Lemma 1.8) in even characteristic.

The **Suzuki curve** is obtained for the values \( q_0 = 2^n \) and \( q = 2^{2n+1} \), see (Hansen and Stichtenoth 1990). So \( q_0 < \sqrt{q} \) and the genus is \( q_0(q - 1) \) by Proposition 2.5.8 which agrees with (Hansen and Stichtenoth 1990, Proposition 1.1).

### 2.6 Classification of trinomial curves

Let \( f(T) \) be a univariate polynomial with coefficients in \( \mathbb{F}_q \). Suppose that \( f(T) \neq 0 \) and that \( f \) consists of the sum of exactly \( d \) monomials. Then we can write

\[ f(T) = \alpha_1 + \alpha_2 T^{m_2} + \cdots + \alpha_d T^{m_d}, \]

where \( \alpha_i \in \mathbb{F}_q^* \) for all \( i \) and \( m_{i-1} < m_i \) for \( 2 \leq i \leq d \) (define \( m_1 = 0 \)). If \( d = 3 \), we call \( f \) a **trinomial**. To obtain a (possibly reducible) plane curve, we use the substitute-and-reduce construction (Construction 2.1.2). First we substitute \( T = X^l Y \). Then we reduce all exponents modulo some number \( v \) to obtain a particular equation \( F_{k,1}(X,Y) \). After this, we can obtain more equations by multiplying \( F_{k,1} \) by \( X^l \) and again reduce all exponents modulo \( v \). We can assume without loss of generality that \( 0 \leq l < v \) and \( m_d < v \). We know that \( Y \) does not divide \( F \), since \( f(0) \neq 0 \). Since we want the equations to be irreducible, we should choose \( l \) such that the resulting \( F \) is not divisible by \( X \). This means that, given \( k \), we should have

\[ l \equiv -km_i \pmod{v} \]

for some \( i \). In this way we obtain \( d \) possible values for \( l \), say \( l_1 \leq l_2 \leq \cdots \leq l_d \). We define \( F_{k,i}(X,Y) \) as the outcome of the above process with the choice \( l = l_i \). Note that \( F_{k,i}(X,Y) \) depends only on the residue class of \( k \) modulo \( v/\gcd(v,m_1,m_2,\ldots,m_d) \). A curve obtained from the substitute-and-reduce construction using a trinomial we call a **trinomial curve**.

**Example 2.6.1** Choose \( f(T) = 1 + T + T^2 \in \mathbb{F}_2[T] \) and \( v = 3 \). Then we obtain:

<table>
<thead>
<tr>
<th>( k )</th>
<th>( F_{k,1}(X,Y) )</th>
<th>( F_{k,2}(X,Y) )</th>
<th>( F_{k,3}(X,Y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( 1 + XY + X^2Y^2 )</td>
<td>( X + X^2Y + Y^2 )</td>
<td>( X^2 + Y + XY^2 )</td>
</tr>
<tr>
<td>2</td>
<td>( 1 + X^2Y + XY^2 )</td>
<td>( X + Y + X^2Y^2 )</td>
<td>( X^2 + XY + Y^2 )</td>
</tr>
<tr>
<td>3</td>
<td>( 1 + Y + Y^2 )</td>
<td>( 1 + Y + Y^2 )</td>
<td>( 1 + Y + Y^2 )</td>
</tr>
</tbody>
</table>
2.6 Classification of trinomial curves

Remark 2.6.2 A generalization of the substitute-and-reduce construction would be to substitute $X^lf(X^nY^m)$ in a polynomial $f(T)$ and then to reduce the exponents modulo $v$. In order to make sure that the same estimate on the number of rational points is obtained, one should demand $\gcd(m,v) = 1$. In that case a positive number $n$ exists such that $v((mn - 1)$. Hence substituting $X^nY^m$ in $f(T)$ gives rise to the same curves as substituting $X^nY$ in $f(T^v)$.

To classify the curves thus obtained up to isomorphism, we introduce the following operation on bivariate polynomials. If $F(X,Y)$ is a bivariate polynomial, then we define $B(F(X,Y)) = X^{-n}F(X,XY)$, where $n$ is the largest integer such that $X^n$ divides $F(X,XY)$. This operator is a special case of blowing up a curve at a point. In particular, the curves defined by the equations $F(X,Y) = 0$ respectively $B(F(X,Y)) = 0$ are birationally equivalent.

To calculate the effect of the operator $B$ on the polynomials $F_{k,l}(X,Y)$ note that

$$B(X^lf(X^nY)) = X^{l-n}f(X^{k+1}Y),$$

where $n$ is the largest natural number such that $X^n$ divides $X^lf(X^{k+1}Y)$. Hence for any $k$ and $i$ some $j$ exists such that $B(F_{k,i}(X,Y))$ and $F_{k+1,i}(X,Y)$ become the same after reducing the exponents of the powers of $X$ of the former modulo $v$. Using this and similar arguments several times and remarking that $F_{v,i}(X,Y) = f(Y)$ for all $i$, we have proved the following proposition.

Proposition 2.6.3 Let the notation be the same as above. For all $k$ and for all $1 \leq i \leq d$, natural numbers $n_1, \ldots, n_d$ exist such that the curves with equations $F_{k,i}(X,Y) = 0$ and $\sum_{j=1}^{d} \alpha_j X^{n_j} Y^{m_j} = 0$ are birationally equivalent.

From the above proposition it is clear that in order to investigate the substitute-and-reduce construction, it suffices to study curves given by the equations

$$\sum_{j=1}^{d} \alpha_j X^{n_j} Y^{m_j} = 0,$$

If $l = \gcd(m_1, m_2, \ldots, m_d) \neq 1$, the above curve is covered by the curve given by the equation

$$\sum_{j=1}^{d} \alpha_j X^{n_j v} Y^{m_j/l} = 0.$$

This covering is in fact a Kummer covering. Since Kummer coverings are well-known and understood, we can restrict ourselves to the case that $l = 1$. Hence, from now on we assume that $\gcd(m_1, m_2, \ldots, m_d) = 1$.

The question remains whether all curves of the above form are mutually distinct or that some of them are birationally equivalent. In fact the latter is the case, since applying the operator $B^v$ preserves the above standard form. More precisely
\[ B^v \left( \sum_{j=1}^{d} \alpha_j X^{n_j v} Y^{m_j} \right) = X^{-N v} \left( \sum_{j=1}^{d} \alpha_j X^{(n_j + m_j) v} Y^{m_j} \right), \]

where \( N \) is some integer.

We will represent from now on the equation \( \sum_{j=1}^{d} \alpha_j X^{n_j v} X^{m_j} \) by the \( d \)-tuple \( \mathbf{n} = (n_1, \ldots, n_d) \). Conversely, we can associate with any \( d \)-tuple of integers \( \mathbf{n} \) the equation \( X^N \sum_{j=1}^{d} \alpha_j X^{n_j v} Y^{m_j} \), where \( N \in \mathbb{Z} \) is chosen uniquely such that \( n_j v + N \geq 0 \) for all \( j \) and \( n_j v + N = 0 \) for some \( l \). For a given \( f(T) \) we find a bijection between the set of curves as in Proposition 2.6.3 and the set \( \mathbb{Z}^d / \langle 1 \rangle \), with \( 1 \) the all one vector. We have shown that

\[ B^v (\mathbf{n} + \langle 1 \rangle) = \mathbf{n} + \mathbf{m} + \langle 1 \rangle, \]

where \( \mathbf{m} = (m_1, m_2, \ldots, m_d) \). Hence, in \( \mathbf{n} + \langle 1, \mathbf{m} \rangle \) any two \( d \)-tuples give rise to mutually birationally equivalent curves. For this reason, we should study the group \( G = \mathbb{Z}^d / \langle 1, \mathbf{m} \rangle \). We can use the theory of finitely generated abelian groups to obtain a simpler form for \( G \). Instead of generating \( \mathbb{Z}^d \) by the standard basis vectors, we can generate \( \mathbb{Z}^d \) by any set of vectors \( \mathbf{v}_1, \ldots, \mathbf{v}_d \) as long as the determinant of the \( d \times d \) matrix formed by these vectors equals \( \pm 1 \). In particular, we can generate \( \mathbb{Z}^d \) by suitable vectors of the form \( 1, \mathbf{m}, \mathbf{a}_1, \ldots, \mathbf{a}_{d-2} \) where \( a_{i1} = 0 \) for all \( i = 1, \ldots, d - 2 \). The determinant condition can be reformulated as

\[
\begin{vmatrix}
    m_2 & \cdots & m_d \\
    a_{12} & \cdots & a_{1d} \\
    \vdots & \ddots & \vdots \\
    a_{d-2 \ 2} & \cdots & a_{d-2 \ d}
\end{vmatrix} = \pm 1,
\]

since \( m_1 = 0 \). Furthermore, we can choose the numbers \( a_{ij} \) all non-negative and assume that the above determinant is \( +1 \). Hence we see that

\[ G \cong \mathbb{Z}^{d-2}. \]

There is one more birational equivalence which we have not taken into account yet, namely the map which assigns to the pair \( (X, Y) \) the pair \( (X^{-1}, Y) \). Hence, using the above notation, the \( d \)-tuples \( \pm \mathbf{n} \) give rise to birationally equivalent curves.

We have proved the following proposition.

**Proposition 2.6.4** Suppose that \( \gcd(m_1, m_2, \ldots, m_d) = 1 \) and let \( \alpha_j \in \mathbb{F}_q^* \) where \( 1 \leq j \leq d \). Any curve with equation

\[ \sum_{j=1}^{d} \alpha_j X^{n_j v} Y^{m_j} = 0 \]

is birationally equivalent to a curve with equation...
\[
\alpha_1 + \sum_{j=2}^{d} \alpha_j X^{m_j} Y^{a_{ij} v} = 0,
\]

where the numbers \(a_{ij}\) are non-negative integers satisfying

\[
\begin{bmatrix}
  m_2 & \cdots & m_d \\
  a_{12} & \cdots & a_{1d} \\
  \vdots & \ddots & \vdots \\
  a_{d-2} & \cdots & a_{d-2} d
\end{bmatrix} = 1
\]

and the \(c_i\) are non-negative integers depending on the \(n_j\).

For \(d = 3\) we find the following corollary.

**Corollary 2.6.5** Suppose that \(\gcd(n,k) = 1\) and let \(\alpha_j \in \mathbb{F}_q^*\) where \(1 \leq j \leq 3\). Any curve with equation

\[
\alpha_1 X^{n_1 v} + \alpha_2 X^{n_2 v} Y^k + \alpha_3 X^{n_3 v} Y^n = 0
\]

is birationally equivalent to a curve with equation

\[
\alpha_1 + \alpha_2 X^{a v} Y^k + \alpha_3 X^{b v} Y^n = 0,
\]

where \(a\) and \(b\) are fixed non-negative integers such that \(kb - an = 1\) and \(c\) is a non-negative integer depending on the numbers \(n_1, n_2\) and \(n_3\).

**Remark 2.6.6** Since \(kb - an = 1\), we can easily determine \(a\) and \(b\) with the Euclidean algorithm. In fact we can determine \(a\) and \(b\) such that the above holds and \(a < b\). The curve is reducible if \(c = 0\) and absolutely irreducible if \(c > 0\). Let \(\mathcal{X}\) be the curve of Corollary 2.6.5 for fixed \(a, b\) and \(v\) and given \(c\). Then \(\mathcal{X}_{c}\) is a covering of \(\mathcal{X}\) under the map \((x, y) \mapsto (x^d, y)\). If \(p\) is the characteristic of the field, then \(\mathcal{X}_{c, p}\) is a purely inseparable covering of \(\mathcal{X}_c\), so they are isomorphic curves. Hence we may assume from now on that \(c\) is not divisible by \(p\).

We assumed before that \(\gcd(m_1, m_2, \ldots, m_d) = 1\), so for a trinomial this means in our notation that \(\gcd(n,k) = 1\). The following theorem however is true as well if \(\gcd(n,k) \neq 1\), just as long \(p\) as does not divide this \(\gcd\). Note that \(p\) does not divide \(v\), since \(v\) is a divisor of \(q - 1\).

**Theorem 2.6.7** Let the curve \(\mathcal{X}_c\) be given by the equation:

\[
\alpha_1 + \alpha_2 X^{a v} Y^k + \alpha_3 X^{b v} Y^n = 0,
\]

where \(c\) is non-zero and not divisible by the characteristic \(p\) and \(kb - an = \gcd(n,k)\). Further suppose that \(p\) does not divide \(\gcd(n,k)\). Finally suppose that \(\alpha_1 \alpha_2 \alpha_3 \neq 0\). Then \(\mathcal{X}\) is an absolutely irreducible curve of genus
\[ g = 1 + \frac{v \gcd(n, k) - \gcd(bv, n) - \gcd(av, k) - \gcd((b - a)v, n - k)}{2}. \]

**Proof:** Denote by \( p \) the characteristic of the field. Using the fact that \( kb - an = \gcd(n, k) \) and the fact that \( p \) does not divide \( cv \gcd(n, k) \), it is easy to show that \( X_c \) does not have any singularities outside the points \([1 : 0 : 0]\) and \([0 : 1 : 0]\).

Now we investigate the question of degeneracy. The Newton polygon is bounded by three edges. One edge connects the points \((0, 0)\) and \((k, acv)\), the second connects the points \((k, avc)\) and \((n, bv)\), while the third connects the points \((n, bv)\) and \((0, 0)\).

We claim that the curve is non-degenerate with respect to its Newton polygon. The question of whether or not the curve is degenerate boils down to the question whether or not the characteristic \( p \) divides one of the numbers \( \gcd(k, avc) \), \( \gcd(n, bv) \), or \( \gcd(n - k, (b - a)cv) \). If \( p \) does not divide either \( k \) or \( n \), this is clear. Since we know that \( p \) does not divide \( \gcd(n, k)cv \), we see that the characteristic cannot divide both \( n \) and \( k \). Suppose that \( p \) divides \( k \) but not \( n \) (the case \( p \) divides \( n \) but not \( k \) is similar). It is immediately clear that \( p \) does not divide either \( \gcd(n, bv) \) or \( \gcd(n - k, (a, b)cv) \). Since \( kb - na = \gcd(n, k) \) we conclude that \( p \) does not divide \( a \). Hence \( p \) does not divide \( \gcd(k, avc) \) either. This proves our claim. Apply Baker's theorem to complete the proof of the theorem. \( \square \)

**Remark 2.6.8** If \( g(X_1) > 1 \), then

\[
\frac{N(X)}{g(X) - 1} \leq \frac{N(X)}{g(X_1) - 1},
\]

where \( N(X) \) is the number of rational points of a curve \( X \) that is defined over a finite field, by (Garcia and Stichtenoth 1996, Lemma 2). Hence from the point of view of counting curves with many rational points relative to its genus, the curve \( X_1 \) is the most interesting one of the sequence. The others can be viewed as Kummer covers of \( X_1 \). If \( c = 1 \) and \( \gcd(n, k) = 1 \), we get the following statement.

**Corollary 2.6.9** Suppose that \( \gcd(n, k) = 1 \) and let \( a < b \) be positive integers such that \( kb - an = 1 \). Also suppose that \( \alpha_1, \alpha_2, \alpha_3 \) are elements of \( \mathbb{F}^* \). Finally let \( v \) be a natural number not divisible by the characteristic of \( \mathbb{F}_q \). The curve given by the equation

\[
\alpha_1 + \alpha_2 X^{av} Y^k + \alpha_3 X^{bv} Y^n = 0,
\]

is absolutely irreducible and has genus

\[
g = 1 + \frac{v - \gcd(v, n) - \gcd(v, k) - \gcd(v, n - k)}{2}.
\]

**Proof:** From the previous theorem we get

\[
g = 1 + \frac{v - \gcd(bv, n) - \gcd(av, k) - \gcd((b - a)v, n - k)}{2}.
\]
2.6 Classification of trinomial curves

Since $bk - an = 1$, we see that we can leave out $b$ from $\gcd(bv, n)$ and likewise the $a$ and $(b - a)$ further on in the formula.

\[\square\]

**Remark 2.6.10** Now we will be more precise about the number of rational points. When we introduced the substitute-and-reduce construction, we remarked that if $f(T)$ is a polynomial of degree less than $v$ having $r$ zeros over the finite field $\mathbb{F}_q$, the construction yields curves with at least $rv$ rational points.

From now on we will assume that $f(T) = \alpha_3 T^n + \alpha_2 T^k + \alpha_1 \in \mathbb{F}_q[T]$, with $\alpha_1 \alpha_2 \alpha_3 \neq 0$. As before we assume that $\gcd(n, k) = 1$. We know from Corollary 2.6.5 that we are really studying curves $\mathcal{X}_c$ given by equations of the form

\[\alpha_1 + \alpha_2 x^{ace} y^k + \alpha_3 x^{bce} y^n = 0,
\]

where $bk - an = 1$ and $a < b$. Indeed this curve has at least $rv$ points, since if we substitute for $X$ some element from $\mathcal{P}$, the equation reduces to $f(Y) = 0$, which has $r$ solutions. However, we did not take into consideration yet the rational points we obtain if we substitute for $X$ an element not from $\mathcal{P}$. For each $x$ from a coset $\xi \mathcal{P}$, the value $x^*$ is constant, say $\beta_\xi$. We choose coset representatives $\xi_1, \ldots, \xi_{(q-1)/v}$ for $\mathcal{P}$ and define $\beta_i \xi = \beta_\xi \xi$. We can assume $\xi_1 = 1$ and hence $\beta_1 = 1$. Now define $r_i$ to be the number of zeros over $\mathbb{F}_q$ of the polynomial $f_i(T) = \alpha_3 \beta_i^{cT} T^n + \alpha_2 \beta_i^{b T^k} + \alpha_1$. Note that $f_i(T) = f(T)$ and hence $r_1 = r$. The number of rational points of the curve $\mathcal{X}_c$ is seen to be at least $v \sum_{i=1}^{(q-1)/e} r_i$. In fact we can state the following proposition. Note that we are still supposing $\gcd(n, k) = 1$ and that the characteristic $p$ does not divide $c$.

**Proposition 2.6.11** The number of rational points $N_1$ of the curve $\mathcal{X}_c$ over the finite field $\mathbb{F}_q$ satisfies the following inequalities:

\[v \sum_{i=1}^{(q-1)/e} r_i \leq N_1 \leq \gcd(k, cv) + \gcd(n, cv) + \gcd(n - k, cv) + v \sum_{i=1}^{(q-1)/e} r_i.
\]

**Proof:** The first inequality is clear from the remark preceding this proposition. Note that the number of affine rational points of $\mathcal{X}_c$ is exactly equal to the lower bound given above. Hence to obtain an upper bound one should look at the points at infinity. These are in general singular points, hence one should study the desingularization tree as done before in detail. Applying Remark 2.3.17 one obtains the result. (Use that since $bk - an = 1$, one can leave out $a, b$ and $b - a$ from the $\gcd$ expressions).

\[\square\]

Note that the $r_i$ do not depend on $c$. Hence the number of rational points of $\mathcal{X}_c$ will not increase significantly (if indeed at all), as $c$ increases, while the genus increases more or less linearly in $c$. This illustrates again that $\mathcal{X}_1$ is by far the most interesting curve of the family as far as the number of rational points relative to the genus is considered.
2.7 Curves with many rational points

In the last section we developed the theory of trinomial curves. In this section we would like to give some examples. First we will show that the trinomial version of the substitute-and-reduce construction gives rise to several well-known curves, such as the Hermitian curve and the Klein quartic as well as some curves mentioned in (Justesen, Larsen, Jensen, Havemose, and Høholdt 1989). After that we will investigate several other curves mentioned in the literature and correct some mistakes. Finally we will give tables of the best results obtained with trinomial curves for small finite fields.

Definition 2.7.1 Let $N_q(g)$ denote the maximal number of $F_q$-rational points of an absolutely irreducible, non-singular projective curve over $F_q$ of genus $g$.

For tables of $N_q(g)$ we refer to (van der Geer and van der Vlugt 2000).

Example 2.7.2 Let $q = 32$ and $f(T) = T^5 + T^2 + 1$ and take $k = 11$ and $l = 9$ in the substitute-and-reduce construction (Construction 2.1.2) as was done in (Justesen, Larsen, Jensen, Havemose, and Høholdt 1989, Example 6). Then we get the plane curve with affine equation

$$X^9 + X^2Y^5 + Y^2 = 0.$$ 

The curve is absolutely irreducible by Proposition 2.2.11 and its genus is 15 by Corollary 2.4.3. This was shown in (Haché and Le Brigand 1995, Example 6.1) where it was noticed that the genus is much smaller than claimed in (Justesen, Larsen, Jensen, Havemose, and Høholdt 1989); it is 15 instead of 26, and the number of rational points is 158 instead of 157. Hence $158 \leq N_{32}(15) \leq 196$, see (van der Geer and van der Vlugt 2000). No better lower bounds for $N_{32}(15)$ are known at the moment.

Let’s study this example in some more detail. Define $f(T) = T^5 + T^2 + 1 \in F_2[T]$. Using Corollary 2.6.5 on the classification of trinomial curves, we only have to look at curves with equation

$$1 + X^2Y^{31c} + X^5Y^{62c} = 0,$$

where $c \in \mathbb{N}_0$. We may assume that $c$ is odd by Remark 2.6.6. We find that the genus $g$ of the non-singular model is

$$g = 1 + \frac{1}{2}(31c - \gcd(c, 2) - \gcd(c, 3) - \gcd(c, 5))$$

by Theorem 2.6.7 and the number $N_1$ of rational points over $F_{32}$ satisfies

$$N_1 \leq 155 + \gcd(c, 2) + \gcd(c, 3) + \gcd(c, 5)$$

by Proposition 2.6.11 and Remark 2.3.17. For $c=1$ we find a curve with $N = 158$ and $g = 15$ which is birationally equivalent to the curve of Example 2.7.2. However,
we can prove that we cannot find a larger value of $N/g$, using this construction and this particular $f(T)$, as was noted in general in Remark 2.6.8. Another idea would be to try to improve this number by using another $f(T)$. However, this will not work if $f(T)$ is a trinomial as a computer search has shown.

**Example 2.7.3** Let $q = 128$ and $f(T) = T^{10} + T + 1$ and take $k = 113$ and $l = 14$ in the substitute-and-reduce construction as in (Justesen, Larsen, Jensen, Havemose, and Høholdt 1989, Example 6). Then we get the plane curve with affine equation $X^{14} + XY^{10} + Y = 0$. The number of rational points is 892. The genus is 63 by Corollary 2.4.3. So $g/N = 0.071$, which is smaller than 0.085 as claimed in (Justesen, Larsen, Jensen, Havemose, and Høholdt 1989). The parameters of this curve are beyond the scope of the tables for $N_q(g)$ in (van der Geer and van der Vlugt 2000).

**Example 2.7.4** Let $f(T) = T^q + T + 1$, $\mathcal{P} = \mathbb{F}_{q^2}^*$, $v = q^2 - 1$, $k = q$ and $l = 0$ in Construction 2.1.2. Then

$$F(X, Y) = 1 + XY^q + X^qY.$$  

So this curve has at least $q(q^2 - 1)$ rational points over $\mathbb{F}_{q^2}$. The corresponding homogeneous polynomial is $F^*(X, Y, Z) = Z^{q+1} + XY^q + X^qY$. Hence there are $q + 1$ rational points at infinity. So, the exact number of $\mathbb{F}_{q^2}$-rational points is $q^2 + 1$. The genus can be computed to be $g = 1 + ((q^2 - 1) - 1 - 1 - (q - 1)) / 2$ which is equal to $q(q - 1)/2$. Therefore the curve is maximal (Stichtenoth 1993). It is in fact the Hermitian curve.

**Example 2.7.5** Again let $f(T) = T^q + T + 1 \in \mathbb{F}_{q^2}[T]$. Suppose $q$ is odd and let $\mathcal{P}$ be the set of non-zero squares. This is a subgroup of $\mathbb{F}_{q^2}^*$ of order $(q^2 - 1)/2$. Then the related curve $X_1$ is given by the equation:

$$X^{(q^2 - 1)/2}Y^q + X^qY + 1 = 0,$$

where $v = (q^2 - 1)/2$. The curve $X_1$ has genus $(q - 1)^2/4$. Using the notation of Remark 2.6.10, we have $f_1(T) = T^q + T + 1$ and $f_2(T) = T^q - T + 1$. Hence $r_1 = q$ and $r_2 = 0$. Further the curve has $q + 1$ points at infinity, which are all rational over $\mathbb{F}_{q^2}$ according to Remark 2.3.17. Hence the curve $X_2$ has $q + 1 + q(q^2 - 1)/2$ rational points over $\mathbb{F}_{q^2}$. Just as the Hermitian curve, this curve is maximal, since its function field can be seen as a subfield of the Hermitian function field. Indeed the Hermitian curve is a Kummer cover of degree 2 of this curve, since the Hermitian curve can also be given by the equation:

$$X^{(q^2 - 1)/2}Y^q + X^qY + 1 = 0,$$

where $v = (q^2 - 1)/2$.

**Example 2.7.6** It is also instructive to examine the Kummer cover of degree 2 of the previous curve given by the equation:
\[
X^{(q^2 + 1)q^2} + X^qY^2 + 1 = 0
\]

We write \( \mathcal{Y} \) for this curve. Using Theorem 2.6.7 one sees that \( \mathcal{Y} \) has genus
\[
g = \frac{q^2 - 3 - (q - 1) \gcd \left( \frac{q^2 + 1}{2}, 2 \right)}{2}.
\]

Now we investigate the number of its rational points. The \( q + 1 \) rational points at infinity of \( \mathcal{X}_1 \) split into \( 4 + (q - 1) \gcd((q + 1)/2, 2) \) rational points of \( \mathcal{Y} \) as can be seen by investigating the local properties of these points. Now we look at the affine points of \( \mathcal{X}_1 \). We know from the above discussion that all \( y \)-coordinates of the affine rational points of \( \mathcal{X}_1 \) are non-zero and are zeros of the polynomial \( f(T) \). Hence to study the number of rational points of \( \mathcal{Y} \), we need to know how many zeros of \( f(T) \) are squares. We will investigate this in the following two lemma.

Lemma 2.7.7 Denote by \( s(\alpha) \), the number of non-zero squares that are a zero of the polynomial \( g(T) = T^q + T + \alpha \in \mathbb{F}_q[T] \) and as before denote by \( v \) the value \( (q^2 - 1)/2 \). Then we have:
\[
s(0) = \begin{cases} 
0 & \text{if } q \equiv 1 \pmod{4}, \\
q - 1 & \text{if } q \equiv -1 \pmod{4}.
\end{cases}
\]

and
\[
s(\alpha) = \frac{v - s(0)}{q - 1},
\]
for all non-zero \( \alpha \in \mathbb{F}_q \).

Proof: First we investigate \( s(0) \). If \( x^q + x = 0 \) and \( x \) is non-zero, we see \( x^{q-1} = -1 \) and hence \( x^{(q^2 - 1)/2} = (-1)^{(q^2 - 1)/2} \). Hence \( s(0) = q - 1 \) if \( q \equiv -1 \pmod{4} \) and \( s(0) = 0 \) if \( q \equiv 1 \pmod{4} \). Now suppose that \( \alpha \in \mathbb{F}_q^* \). Further denote by \( \rho \) a primitive root of \( \mathbb{F}_q \). If \( x \) is a zero of \( T^q + T + \alpha \), then \( x \rho \) is a zero of \( T^q + T + \alpha \rho \). Since \( \rho \) is a square in \( \mathbb{F}_{q^2} \) (indeed all elements of \( \mathbb{F}_q \) are squares in \( \mathbb{F}_{q^2} \)) we see \( s(\alpha) = s(\alpha \rho) \). Hence all \( s(\alpha) \) are the same for \( \alpha \) non-zero. Since any square is a zero of \( T^q + T + \alpha \) for some \( \alpha \), we have \( \sum_{\alpha \in \mathbb{F}_q} s(\alpha) = (q^2 - 1)/2 \). Hence we obtain the desired result.

We see that the number of \( \mathbb{F}_{q^2} \)-rational points of \( \mathcal{Y} \) is
\[
2s(1)v + 4 + (q - 1) \gcd \left( \frac{q + 1}{2}, 2 \right).
\]

We state the above results in the following proposition.

Proposition 2.7.8 Let \( q \) be a power of an odd prime and denote by \( v \) the number \( (q^2 - 1)/2 \). The curve given by the equation:
\[
X^{(q^2 + 1)q^2} + X^qY^2 + 1 = 0,
\]
is absolutely irreducible and has genus
\[ g = \begin{cases} \frac{q^2 - q - 1}{2} & \text{if } q \equiv 1 \pmod{4}, \\ \frac{q^2 - 2q - 1}{2} & \text{if } q \equiv -1 \pmod{4}. \end{cases} \]

Further the number of \( F_q \)-rational points is given by
\[ N_1 = \begin{cases} (q + 1) \frac{q^2 + 1}{2} + 2 & \text{if } q \equiv 1 \pmod{4}, \\ (q - 1) \frac{q^2 + 3}{2} + 4 & \text{if } q \equiv -1 \pmod{4}. \end{cases} \]

For \( q = 3 \) we find a curve of genus 1 having 16 points defined over \( F_3 \), which is maximal. For \( q = 9 \) we find a curve of genus 35 having 412 points over \( F_{31} \). At the moment the tables in (van der Geer and van der Vlugt 2000) do not contain an entry for this particular genus and field size. Using the Oesterlé upper bound we have proved \( 412 \leq N_{31}(35) \leq 711 \).

**Example 2.7.9** Let \( g(T) = \alpha_2 T^q - 1 + \alpha_1 T^{q-1} + \alpha_0 \), where the \( \alpha_i \) are non-zero elements of \( F_q \). Then \( g(T) \) has \( q^2 - 1 \) distinct zeros in \( F_q \) if and only if \( (\alpha_1/\alpha_0)q^2 + q + 1 = 1 \) and \( \alpha_2 = (\alpha_1/\alpha_0)^{-q} \alpha_0 \). Use \( g(T), P = F_q^*, v = q^3 - 1, k = q \) and \( l = q - 1 \) with \( q^2 - 1 \) zeros in \( F_q \) in the substitute-and-reduce construction. Then
\[ F(X,Y) = \alpha_0 X^{q-1} + \alpha_1 X^{q^2-1} Y^{q-1} + \alpha_2 Y^{q^2-1}. \]

In this way, we get a curve with at least \( (q^2 - 1)(q^3 - 1) \) rational points over \( F_q \). The plane curve has three singular points and \( q - 1 \) branches meet at every singular point. The main part of the polynomial at such a singular point has one edge \( \gamma \) from \( (q - 1,0) \) to \( (0,q^2 - 1) \), and the corresponding univariate polynomial is \( f_{\gamma}(T) = \alpha_i T^{q^2 - 1} + \alpha_j \). If \( \alpha_i = 1 \) for all \( i \), then \( f_{\gamma}(T) \) has \( q - 1 \) zeros in \( F_q \) if \( q \) is even. So the branches are \( F_q \)-rational in even characteristic, by Remark 2.3.17. On the other hand if the characteristic is odd, then at all singular points the branches are not rational. We conclude that the exact number of points of the non-singular model of this curve is \( (q^2 - 1)(q^3 - 1) + 3(q - 1) \) in even characteristic and \( (q^2 - 1)(q^3 - 1) \) in odd characteristic.

This example is a specialization of Example 2.4.5 with \( m = q^2 - 1 \) and \( n = q - 1 \). So \( m^2 - mn + n^2 = (q - 1)^2(q^2 + q + 1) \) and is relatively prime to the characteristic. Hence the genus is
\[ g = 1 + \frac{1}{2} \left( (q - 1)^2(q^2 + q + 1) - 3(q - 1) \right). \]

In (Justesen, Larsen, Jensen, Havemose, and Hoholdt 1989, Example 5) these curves were mentioned, but a formula for the genus was not given.

If \( q = 3 \) we get a curve of genus 24 having 208 rational points over \( F_{27} \). The best-known upper bound is \( N_{27}(24) \leq 235 \). In (Shbat 1979) 190 \( \leq N_{27}(24) \) is shown and 208 \( \leq N_{27}(24) \) is proved in (van der Geer and van der Vlugt 1999) both by means of Kummer coverings.
Example 2.7.10 A generalization of the Klein quartic is obtained by the reduction of \( Xf(X^vY) \), where \( f(T) = T^{q+1} + T + 1 \) and \( P \) is the cyclic subgroup of \( \mathbb{F}_q^* \) of order \( v = q^2 + q + 1 \). Then the curve with defining equation

\[
X^q Y + Y^q + X = 0
\]

has at least \((q + 1)(q^2 + q + 1)\) rational points over \( \mathbb{F}_q \). This example has been studied for curves representing designs in (Pellikaan 1998). There it was shown that the exact number of rational points is equal to \( 2q^3 + 1 + (1 - \varepsilon_q)(q^2 + q + 1) \), where \( \varepsilon_q \) is the remainder of \( q + 1 \) modulo 3. This example is a specialization of Example 2.4.5 with \( m = q + 1 \) and \( n = 1 \). So \( m^2 - mn + n^2 = q^2 + q + 1 \) and this value is relatively prime to the characteristic. Hence the genus is \( g = q(q + 1)/2 \).

Remark 2.7.11 Let \( f(T) \in \mathbb{F}_q[T] \) be a polynomial with \( r \) zeros in \( \mathbb{F}_q^* \). Let \( s \) be a positive integer dividing \((q^r - 1)/(q - 1)\). Then \( f(T^s) \) has at least \( rs \) zeros in \( \mathbb{F}_q^* \).

Example 2.7.12 The polynomial \( f(T) = T^q + T - 1 \) has \( q \) zeros in \( \mathbb{F}_{q^2} \). Let \( s = q^2 + 1 \). Then \( f(T^s) \) has \( q^2 + q \) zeros in \( \mathbb{F}_q \) by Remark 2.7.11. Apply Construction 2.1.2 to \( f(T^s) \) with \( P = \mathbb{F}_q^* \), \( k = q \) and \( l = 1 \). Then we get

\[
F(X, Y) = X^{q^2} Y^q + qX^{q^2} + qY^{q^2} + 1.
\]

Our methods give that the corresponding non-singular curve has genus \( g \) and \( N \) rational points, where

\[
g = \frac{1}{2} q(q^5 + q^3 - q^2 - 2q - 1) \quad \text{and} \quad N = (q^2 + 1)(q^5 + 1).
\]

This results is obtained in (García and Stichtenoth 1999, Example 4.4) by other methods. The homogeneous polynomial of \( F(X, Y) \) is

\[
F^*(X, Y, Z) = X^{2q^2} Y^{q^3} + qZ + X^{q^3 + q} Y^{q^2 + 1} - Z^{q^3 + q^2 + q^4 + 1}.
\]

So \( F^*(1, Y, Z) = Y^{q^3 + q} + Y^{q^2 + 1} - Z^{q^3 + q^2 + q^4 + 1} = 0 \) gives another plane model of the curve, the one found by (García and Stichtenoth 1999). There it was noted that for \( q = 2 \) we get a curve of genus 31 having 165 rational points over \( \mathbb{F}_{16} \), which is an improvement on the previously-known lower bound for \( N_{16}(31) \).

Example 2.7.13 In (Moreno, Zinoviev, and Zinoviev 1995), bivariate binary polynomials are investigated to find curves with many rational points over finite fields of characteristic 2 of genus 3, 4 and 5. The genus is determined by computing the delta-invariant at the singular points. Also Theorem 2.4.2 on the Newton polygon can be applied. Consider for instance

\[
F(X, Y) = X^2 + XY + Y^2 + X^3 + Y^3 + X^5 + X^3Y^2 + X^2Y^3
\]

of (Moreno, Zinoviev, and Zinoviev 1995, Statement 1.1). The Newton polygon is not degenerate and has 4 interior integral points. Moreover \((0 : 0 : 1)\) and \((0 : 1 : 0)\)
are the only singular points. Hence the genus of this curve is 4. The number of rational points over $F_8$ of the non-singular model is 25. The best-known bounds are $25 \leq N_8(4) \leq 27$ by (van der Geer and van der Vlugt 2000). In (Moreno, Zinoviev, and Zinoviev 1995, Statement 1.2) there are two mistakes. It is claimed that the non-singular model of the curve with equation
\[
X^2 + XY + Y^2 + X^3 + X^2Y + XY^2 + X^2Y^2 + Y^4 + X^5 = 0
\]
has genus 5 and 26 rational points over $F_8$. But there is another node at $(0,1)$ with two branches that are not rational over $F_8$. So the non-singular model has genus 4 and 25 rational points. Further it is claimed that the non-singular model of the curve with equation
\[
X^2 + X^3 + X^2Y + XY^2 + X^3Y + Y^4 + X^5 + X^4Y + X^3Y^2 = 0
\]
has genus 5 and 26 rational points over $F_8$, however the Newton polygon indicates that the genus is 4. The singular point is not a cusp but after one blow-up we get an ordinary double point with two branches that are not rational. So, the non-singular model has genus 4 and 25 rational points. In (Moreno, Zinoviev, and Zinoviev 1995) no proof is given of the fact that the polynomials considered there are absolutely irreducible. However, this can be shown using Proposition 2.2.15, since all curves have degree 5 and $\sum \delta_p < 4$ in all cases.

**Example 2.7.14** Consider again the curve of Proposition 2.5.8 with equation
\[
X^{q^2} - Y^{q^{q+1}} + X^{q^{-1}}Y^{q+1} - X^{q+q^{-1}} = 0,
\]
over $F_q$, where $q_0$ and $q$ are powers of the prime $p$. The non-singular model has $q^2 + 1$ rational points over $F_q$. The genus is $q_0(q-1)/2$ if $q_0 > \sqrt{q}$, and $q(q-1)/2q_0$ if $q_0 < \sqrt{q}$. For given $q$ the smallest genus is obtained by
\[
g = \begin{cases} 
\frac{1}{2}p\sqrt{q}(q-1) & \text{if } q \text{ is a square}, \\
\frac{1}{2}\sqrt{pq}(q-1) & \text{if } q \text{ is not a square}.
\end{cases}
\]
In the case that $p = 2$ and $q$ is an odd power of 2 we get the Suzuki curve, as noticed in Remark 2.5.9, and the number of rational points of these curves is optimal for given genus by (Hansen and Stichtenoth 1990, Proposition 2.1). For $q = 4$ we get a curve of genus 6 with 17 rational points, which is smaller than $N_4(6) = 20$. For $q = 3$ we get a curve of genus 3 with $N_3(3) = 10$ rational points. For $q = 9$ we get a curve of genus 36 with 82 rational points, while 118 $\leq N_9(36) \leq 142$ by (van der Geer and van der Vlugt 2000). For $q = 27$ we get a curve of genus 117 with 730 rational points. This is beyond the scope of the existing tables. The Oesterlé bound for gives $N_{27}(117) \leq 859$.

It is reasonable to conjecture that $N_q(g)$ is (strictly) increasing in $q$. The best-known lower bound for $N_{27}(124)$ is 680 by (van der Geer and van der Vlugt 1999). The conjecture would imply that $730 \leq N_{27}(117) \leq N_{27}(124)$.
Now we give some tables with the results obtained using trinomials in the substitute-and-reduce construction. We denote the size of the field by $q$. These curves were found by doing a computer search over trinomials $f(T)$ of degree less than $v$. Nearly all curves found will have an equation of the form $1 + X^{av}Y^k + \alpha X^{bv}Y^v$, where $a$ and $b$ are as in Theorem 2.6.7. We will denote this information by giving in the columns after the number of rational points first the value of $n$, $k$ and $v$ and then the minimal polynomial $m_{\alpha}(t)$ of $\alpha$. However, if $\alpha = 1$, we will not type the polynomial $t - 1$, but just leave it out. The only cases when the curve does not have an equation of the form mentioned above are:

- $q = 16, v = q - 1$ and $g = 4$, the curve is given by the equation
  \[ X^{9v}Y^{11} + \alpha X^{4v}Y^5 + 1 = 0, \]
  where $m_{\alpha}(t) = t^4 + t^3 + t$
- $q = 16, v = q - 1$ and $g = 19$, the curve is given by the equation
  \[ X^{3v}Y^{6} + \alpha X^{v}Y^{3} + 1 = 0, \]
  where $m_{\alpha}(t) = t^2 + t + 1$

The column below the $N_1$ has entries of the form $a - b$. The $a$ corresponds to the best-known lower bound for $N_q(g)$ using only trinomials, while the $b$ denotes the best-known upper bound for $N_q(g)$. This upper bound will in this table be either the upper bound mentioned in (van der Geer and van der Vlugt 2000) or the Oesterlé upper bound. In the case $a = b$, as happens for example in the case of the Hermitian curve, we will put $a$ in the table rather than $a - a$.

<table>
<thead>
<tr>
<th>$q = 16$</th>
<th>$q = 32$</th>
<th>$q = 64$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g$</td>
<td>$N_1$</td>
<td>$n,k,v,m_\alpha$</td>
</tr>
<tr>
<td>4</td>
<td>36 - 46</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>37 - 54</td>
<td>5, 1, 15</td>
</tr>
<tr>
<td>6</td>
<td>65</td>
<td>4, 1, 15</td>
</tr>
<tr>
<td>7</td>
<td>33 - 70</td>
<td>2, 1, 15</td>
</tr>
<tr>
<td>19</td>
<td>93 - 134</td>
<td>-</td>
</tr>
<tr>
<td>31</td>
<td>165 - 197</td>
<td>10, 5, 15</td>
</tr>
<tr>
<td>21</td>
<td>191 - 400</td>
<td>37, 16, 63</td>
</tr>
<tr>
<td>24</td>
<td>136 - 446</td>
<td>37, 9, 63</td>
</tr>
<tr>
<td>27</td>
<td>263 - 494</td>
<td>18, 1, 63, $t^2 + t + 1$</td>
</tr>
<tr>
<td>28</td>
<td>513</td>
<td>8, 1, 63</td>
</tr>
<tr>
<td>30</td>
<td>383 - 536</td>
<td>6, 1, 63</td>
</tr>
<tr>
<td>31</td>
<td>318 - 547</td>
<td>31, 17, 63</td>
</tr>
</tbody>
</table>
### 2.7 Curves with many rational points

<table>
<thead>
<tr>
<th>$q = 27$</th>
<th></th>
<th></th>
<th>$q = 81$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$g$</td>
<td>$N_1$</td>
<td>$n, k, v, m_\alpha$</td>
<td>$g$</td>
<td>$N_1$</td>
<td>$n, k, v, m_\alpha$</td>
</tr>
<tr>
<td>6</td>
<td>55 - 86</td>
<td>4, 1, 13</td>
<td>10</td>
<td>182 - 262</td>
<td>20, 1, 40, $t^4 + t^2 - 1$</td>
</tr>
<tr>
<td>12</td>
<td>106 - 146</td>
<td>4, 1, 26</td>
<td>14</td>
<td>169 - 334</td>
<td>16, 1, 40, $t^4 + t^3 + t^2 - t - 1$</td>
</tr>
<tr>
<td>24</td>
<td>208 - 235</td>
<td>8, 2, 26</td>
<td>15</td>
<td>172 - 352</td>
<td>10, 1, 40, $t^4 - t^3 - 1$</td>
</tr>
<tr>
<td>16</td>
<td>370</td>
<td>9, 1, 40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td><strong>388</strong> - 388</td>
<td>15, 1, 40</td>
<td>18</td>
<td>246 - 406</td>
<td>12, 1, 40, $t^2 + 1$</td>
</tr>
<tr>
<td>19</td>
<td>244 - 424</td>
<td>7, 1, 40, $t^2 + t - 1$</td>
<td>20</td>
<td>162 - 442</td>
<td>40, 1, 80, $t^4 - t - 1$</td>
</tr>
<tr>
<td>29</td>
<td>264 - 603</td>
<td>20, 2, 40</td>
<td>30</td>
<td>262 - 621</td>
<td>20, 1, 80, $t^4 - t - 1$</td>
</tr>
<tr>
<td>31</td>
<td>172 - 638</td>
<td>10, 2, 40, $t^4 + t^2 - t + 1$</td>
<td>32</td>
<td>402 - 657</td>
<td>32, 1, 80</td>
</tr>
<tr>
<td>33</td>
<td>256 - 675</td>
<td>12, 2, 40</td>
<td>34</td>
<td>246 - 693</td>
<td>25, 1, 80, $t^4 + t^3 + t^2 - t - 1$</td>
</tr>
<tr>
<td>35</td>
<td>412 - 711</td>
<td>18, 2, 40</td>
<td>36</td>
<td>730</td>
<td>9, 1, 80</td>
</tr>
<tr>
<td>37</td>
<td>488 - 743</td>
<td>15, 1, 80, $t + 1$</td>
<td>38</td>
<td>406 - 756</td>
<td>13, 1, 80</td>
</tr>
<tr>
<td>39</td>
<td>404 - 769</td>
<td>14, 1, 80, $t^2 + t - 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Remark 2.7.15** In (van der Geer and van der Vlugt 2000) a lower bound is mentioned if it is larger than the integral part of $1/\sqrt{2}$ times the Ihara upper bound. Most of the lower bounds mentioned above are already known or do not meet the mentioned criterion, except the case $q = 17$ and $q = 81$. Hence we see $288 \leq N_{81}(17) \leq 388$. This is the reason that we put 288 in boldface.
Chapter 3

Semigroups

Summary In this chapter we will investigate the problem of how to compute the Weierstrass semigroup of a point. As in the last chapter, the theory of Newton polygons will play an important role. Further, we will make use of regular differential forms.

3.1 Introduction

The main object of interest in this chapter is the Weierstrass semigroup of a point on an algebraic curve. Most of the time we will simply speak about the semigroup of a point and delete the prefix “Weierstrass”.

Definition 3.1.1 Let $\mathcal{X}$ be an absolutely irreducible algebraic curve and let $P$ be a point of $\mathcal{X}$. The valuation $v_P$ at $P$ is a map which assigns to an algebraic function on $\mathcal{X}$ its zero order at $P$.

Definition 3.1.2 Let $P$ be a point on a curve $\mathcal{X}$. Denote by $K_\infty(P)$ the ring of algebraic functions on $\mathcal{X}$ which are regular outside $P$. The (Weierstrass)semigroup $W_P$ of the point $P$ is the semigroup of pole orders of functions in $K_\infty(P)$, i.e.

$$W_P = \{-v_P(f) \mid f \in K_\infty(P)\}.$$

Example 3.1.3 Consider the Klein quartic $\mathcal{X}$ given by the equation

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

over the field $\mathbb{F}_2$. Denote by $P$ the point $[1 : 0 : 0]$ and by $Q$ the point $[0 : 1 : 0]$, the only points at infinity. We want to compute the semigroup of $P$. First we will determine the ring of functions which are regular everywhere outside $P$. Denote by $x$ (respectively $y$) the residue classes of $X$ (respectively $Y$) in the ring $\mathbb{F}_2[X,Y]/(X^3Y + Y^3 + X)$. So, we use the capitals $X$ and $Y$ to denote the coordinates
and \(x\) and \(y\) to denote the corresponding coordinate functions. It is clear that \(x\) and \(y\) are functions which are regular outside \(P\) and \(Q\). Hence, any monomial \(x^i y^j\) is a function that is regular outside \(P\) and \(Q\). As a matter of fact, looking at the equation of \(X\), one sees that \(x^i y^j\) is a regular function outside \(P\) and \(Q\) as long as \(3i + j \geq 0\). For example, the function \(x/y^3\) is regular outside \(P\) and \(Q\) since
\[
x/y^3 = 1 + x^3/y^2 = 1 + y(x^3 + y^2)^3.
\]
Of course we are not done yet, since some of these functions may have poles at \(Q\) as well. So now we investigate their behaviour in \(Q\). The homogenized equation of the Klein quartic is
\[
X^3 Y + Y^3 Z + X Z^3 = 0
\]
The homogenized version of the function \(x\) (respectively \(y\)) is \((x/z)\) (respectively \((y/z)\)). We get an affine equation where \(Q\) is the point \((0, 0)\) by putting \(Y = 1\) in the homogenized equation. The result is
\[
X^3 + Z + X Z^3 = 0.
\]
Our original functions \(x\) and \(y\) become \(x/z\) and \(1/z\) in the new representation. Denote by \(v_Q\) the valuation at \(Q\). Using the above equation we get the functional functions \(x^3 + x + x z^3 = 0\). Since \(x\) and \(z\) have positive valuation at \(Q\), we see that
\[
v_Q(z + x z^3) = v_Q(z)
\]
and hence
\[
3v_Q(x) = v_Q(x^3) = v_Q(z).
\]
In fact, one can show that \(v_Q(x) = 1\) and hence \(v_Q(z) = 3\). Using this we see (using the homogenized version of the original functions \(x\) and \(y\)) that
\[
v_Q(x/z) = -2
\]
and
\[
v_Q(y/z) = -3.
\]
Hence the function \((x/z)^i (y/z)^j\) is regular at \(Q\) if and only if \(-2i - 3j \geq 0\).

We have found regular functions of the form \(x^i y^j\) (in the original notation). We encountered some restrictions on the \(i\) and \(j\) namely \(3i + j \geq 0\) and \(-2i - 3j \geq 0\). One can show that the functions satisfying these restrictions generate \(K_\infty(P)\). We want to know the pole orders at \(P\) of these functions. In a similar manner one can show that \(v_P(x/z) = -1\) and \(v_P(y/z) = 2\). Hence, we find as semigroup
\[
\{i - 2j \mid 3i + j \geq 0, -2i - 3j \geq 0\},
\]
which equals
\[
\{0, 3, 5, 6, 7, 8, \ldots \}.
\]
3.2 Gaps and differential forms

In the previous example we calculated the semigroup $W_P$ by first calculating the ring $K_\infty(P)$. In general, it is a very difficult problem to calculate this ring explicitly. In the next paragraph we discuss a method of calculating the semigroup $W_P$ without first calculating the ring $K_\infty(P)$. We will apply this theory to calculate the semigroups of plane curves of type II. Further, some results will be presented concerning the semigroups of the curves occurring in the first Garcia-Stichtenoth tower.

3.2 Gaps and differential forms

From now on we suppose that $F$, the field of definition of the curve, is perfect. Hence the case of an algebraic closed field or a finite field is covered.

**Definition 3.2.1** Let $\mathcal{X}$ be an algebraic curve and let $P$ be a point of $\mathcal{X}$. An element from the set $\mathbb{N}_0 \setminus W_P$ is called a gap at $P$.

The following theorem by Weierstrass relates the number of gaps with the genus of the curve.

**Theorem 3.2.2** Let $\mathcal{X}$ be an algebraic curve over a field $F$ of genus $g$ and let $P$ be a point of $\mathcal{X}$. Then the number of gaps at $P$ is equal to $g$.

**Proof:** Denote as before by $L(iP)$ the vector space of functions $f$ in $K_\infty(P)$ such that $f$ has a pole of order at most $i$ at $P$. Consider the dimension sequence $(\dim(L(iP)))_{i \geq 0}$. According to Riemann’s theorem (see Theorem 1.1.2),

$$\dim(L(iP)) = i - g + 1$$

for $i > 2g - 2$. Further $\dim(L(0)) = 1$, since $L(0) = F$. It is also clear that the dimension sequence is non-decreasing. Finally, from the properties of the $L(iP)$ we see that

$$\dim(L((i + 1)P)) \leq \dim(L(iP))$$

We conclude from these facts that

$$\dim(L((i + 1)P)) = \dim(L(iP))$$

for exactly $g$ different values of $i$. However, this equality is equivalent to the statement that $i + 1$ is a gap at $P$. \( \square \)

**Corollary 3.2.3** Let the notation be as above. Let $i + 1$ be a gap at $P$, then

$$i \leq 2g - 2.$$
Proof: This follows from the proof of the above theorem by remarking that
\[
\dim(L((i + 1)P)) = \dim(L(iP)) + 1
\]
whenever \(i > 2g - 2\). \(\square\)

Remark 3.2.4 This theorem gives a way to check whether or not one has found a basis of the ring \(K_\infty(P)\) of functions regular outside \(P\). All one has to do is calculate the pole orders of the basis elements and see whether or not the set of pole orders has \(g\) gaps. For example in Example 3.1.3 the set of pole orders was \(\{0, 3, 5, 6, \ldots\}\), while the genus of the Klein quartic is 3 (as can be seen from, for example, Baker’s theorem). Hence, we have found in that example a basis for all of \(K_\infty(P)\) and in particular we have calculated the semigroup \(W_P\).

Another way of calculating the semigroup of a point is by looking at regular differential forms. The following proposition explains the connection between gaps of a semigroup and regular differential forms.

Proposition 3.2.5 Let \(X\) be an algebraic curve of genus \(g\) defined over the field \(F\). Let \(P\) be a point of \(X\). Let \(\omega\) be a regular differential form on \(X\). Then \(v_P(\omega) + 1\) is a gap at \(P\).

Proof: Denote by \(\Omega(iP)\) the linear space of regular differential forms which have a zero of order at least \(i\) at \(P\). Let \(L(P)\) be as before, Serre duality and Riemann-Roch’s theorem (see Chapter I, Section 5 of (Stichtenoth 1993) ) give
\[
\dim(L(iP)) - \dim(\Omega(iP)) - i = 1 - g.
\]
Note that the right-hand side does not depend on \(i\). We know that \(i + 1\) is a gap at \(P\) if and only if
\[
\dim(L((i + 1)P)) = \dim(L(iP)),
\]
Using the above identity we conclude that \(i + 1\) is a gap if and only if
\[
\dim(\Omega(iP)) - \dim(\Omega((i + 1)P)) = 1,
\]
i.e. if and only if there exist a regular differential of zero order \(i\) at \(P\). \(\square\)

We now have another way of finding a semigroup of a point \(P\). Instead of calculating \(K_\infty(P)\) as in Example 3.1.3 we can calculate \(\Omega(0)\). This may be easier in some cases.

Example 3.2.6 Consider again the Klein quartic. As before we denote by \(P\) the point \([1 : 0 : 0]\) and by \(Q\) the point \([0 : 1 : 0]\). Define
\[
F(X, Y) = X^3Y + Y^3 + X.
\]
The space of regular differential forms is generated by
3.2 Gaps and differential forms

\[
\frac{dx}{x^3 + y^2}, \quad \frac{x \, dx}{x^3 + y^2} \text{ and } \frac{y \, dx}{x^3 + y^2}.
\]

We will prove this in Theorem 3.3.2. Note that we use small characters \( x \) and \( y \) for functions and capitals \( X \) and \( Y \) for coordinates. Assuming this result for now, let's calculate the valuations at \( P \). Remember that \( v_P(x) = -1 \) and \( v_P(y) = 2 \). Hence \( v_P(x^3 + y^2) = -3 \) and \( v_P(dx) = -2 \). Using this we find that the gaps are given by the numbers 2, 1 and 4. This is in agreement with Example 3.1.3.

Now that we know that gaps correspond to regular differential forms, we can prove some more facts about gaps. In the following proposition suppose that we have two curves and an algebraic map between them. The idea that we can lift regular differentials from one curve to the other gives rise to the following proposition. We will first give a definition.

**Definition 3.2.7** Let \( \phi : \mathcal{X} \to \mathcal{Y} \) be a non-constant map of regular algebraic curves. Let \( Q \) be a point of \( \mathcal{X} \) and write \( P = \phi(Q) \). Denote by \( \phi^* : \mathcal{F}(\mathcal{Y}) \to \mathcal{F}(\mathcal{X}) \) the inclusion of function fields induced by \( \phi \). The **ramification index** of \( \phi \) at \( P \), denoted by \( e_{Q|P} \), is defined by the functional equation

\[
e_{Q|P} v_P = v_Q.
\]

Of course it is not immediately clear that the ramification can be defined in this way. For more details see p. 28 of (Silverman 1986).

**Proposition 3.2.8** Let \( \mathcal{X} \) and \( \mathcal{Y} \) be regular algebraic curves. Suppose that we have a non-constant and separable map

\[
\phi : \mathcal{X} \to \mathcal{Y}.
\]

Let \( P \) be a point of \( \mathcal{Y} \) and suppose that \( Q \in \phi^{-1}(P) \). Denote by \( s \) (respectively \( t \)) a uniformizing parameter for \( P \) (respectively \( Q \)) and by \( e_{Q|P} \) the ramification index. If \( a \) is a gap at \( P \), then

\[
(a - 1)e_{Q|P} + v_Q \left( \frac{ds}{dt} \right) + 1
\]

is a gap at \( Q \).

**Proof:** It is convenient to prove this proposition in the language of function fields. We have an inclusion of function fields \( K(\mathcal{Y}) \subset K(\mathcal{X}) \). Denote by \( \Omega_X \) (respectively \( \Omega_Y \)) the space of differentials on \( \mathcal{X} \) (respectively \( \mathcal{Y} \)). Since \( \phi \) is a non-constant and separable map, it induces an injection from \( \Omega_Y \) to \( \Omega_X \) (see for example p.35 of (Silverman 1986)).

We have seen that \( a \) is a gap at \( P \) if and only if there exists a regular differential form \( \omega \) such that \( v_P(\omega) = a - 1 \). Since \( s \) is a uniformizing parameter for \( P \), we can write
\[ \omega = f(s)ds \]
for some Laurent series \( f \). Since \( f(s)ds \) is a regular differential form on \( \mathcal{X} \), when interpreted as a differential form on \( \mathcal{X} \), we conclude that \( v_Q(f(s)ds) + 1 \) is a gap at \( Q \). However, since \( v_Q = e_{Q|P}v_P \), we conclude
\[
v_Q(f(s)ds) = e_{Q|P}(a - 1) + v_Q \left( \frac{ds}{dt} \right).
\]
This concludes the proof. \( \square \)

**Corollary 3.2.9** Let the notation be the same as above. Suppose that the characteristic \( p \) does not divide \( e_{Q|P} \). If \( a \) is a gap at \( P \), then \( ae_{Q|P} \) is a gap at \( Q \). If \( p \) does divide \( e_{Q|P} \) then \( Q \) has a gap strictly larger than \( ae_{Q|P} \).

**Proof:** We know that in \( K(\mathcal{X}) \) the following holds
\[
s \equiv t^{e_{Q|P}} \pmod{t^{e_{Q|P} + 1}}.
\]
Hence
\[
\frac{ds}{dt} \equiv e_{Q|P}t^{e_{Q|P} - 1} \pmod{t^{e_{Q|P}}},
\]
However, if the characteristic does not divide \( e_{Q|P} \), this implies that
\[
v_Q \left( \frac{ds}{dt} \right) = e_{Q|P} - 1,
\]
while if it does divide \( e_{Q|P} \), we see
\[
v_Q \left( \frac{ds}{dt} \right) > e_{Q|P} - 1.
\]
This proposition enables us to find information about semigroups using coverings of curves.

**Example 3.2.10** Let \( F \) be an algebraically closed field. Suppose that we have a non-constant separable algebraic map between elliptic curves defined over \( F \). Since an elliptic curve has genus one, any point on it has semigroup \( \{0, 2, 3, \ldots \} \). Hence the only gap is 1. Using Corollary 3.2.9, we see that all \( e_{Q|P} \) are one. Hence all non-constant separable maps between elliptic curves are unramified. This can also be seen using Hurwitz’s genus formula.
3.3 Semigroups of plane curves

In this section we will focus on the case that the algebraic curve $\mathcal{X}$ is given by one polynomial equation in two variables. As mentioned before the curve is called a plane curve in this case. It is true that any algebraic curve over a perfect field can always be given by one equation in two variables, but sometimes such an equation is very difficult to find. We have seen in the previous section that regular differential forms are connected with gaps. The following theorem (which is contained in [Hovanskii 1978] if the field of definition is $\mathbb{C}$) gives all regular differential forms in a particular case. Before we prove it however, we state a proposition that will be useful in the proof of the theorem. We saw in Example 3.2.6 that to compute the gaps from differential forms, one needs to know the valuations of $x$ and $y$ for the point one is interested in. Note that $x$ (respectively $y$) stands for the image of $X$ (respectively $Y$) in $\mathbb{F}[X,Y]/(F(X,Y))$. The proposition deals with this problem.

**Proposition 3.3.1** Let $\mathcal{X}$ be an absolutely irreducible curve over a perfect field $\mathbb{F}$ given by the plane equation $F(X,Y) = 0$. Suppose that $F$ is non-degenerate with respect to its Newton-polygon. Denote by $P$ an infinitely near point of $[0:0:1]$, $[0:1:0]$ or $[1:0:0]$ and assume that the edge $\gamma$ corresponding to it is given by the equation

$$ai + bj = c,$$

where $a$ and $b$ are integers such that $\gcd(a,b) = 1$. Further assume that the Newton diagram of $F$ is contained in the half space

$$ai + bj \geq c.$$

Then

$$v_P(x) = a$$

and

$$v_P(y) = b.$$

A uniformizing parameter $t$ for $P$ is given by

$$t = x^{-B}y^{A},$$

where $A$ and $B$ are the smallest non-negative integers such that $bA - aB = 1$.

**Proof:** We will prove the case that $P$ is infinitely near to $[0:0:1]$. The other cases are similar. First we will isolate the edge $\gamma$ as we did in Lemma 2.5.3 and after that we will investigate the equation we have obtained. First we suppose that the edge corresponds to the lower part of the Newton polygon, since the general case can be done in a similar way. So, let $\gamma$ be an edge of the Newton diagram
of $F$ connecting $(e_1, f_1)$ and $(e_2, f_2)$. Define as before the length of $\gamma$, denoted by $l(\gamma)$, to be the positive integer $\gcd(e_2 - e_1, f_1 - f_2)$. Then we know that, since $F$ is non-degenerate with respect to its Newton polygon, there correspond $l(\gamma)$ points with $\gamma$ on the non-singular model. According to Lemma 2.5.3, the map

$$\phi(X, Y) = (X^b Y^{-a}, X^{-b} Y^A)$$

is the map that isolates $\gamma$. Define

$$(U, V) = (X^b Y^{-a}, X^{-b} Y^A),$$

The map $\phi$ gives a birational map between the curve $\mathcal{X}$, which is given by the equation $F(X, Y) = 0$ and another curve $\mathcal{Y}$ given by some equation $G(U, V) = 0$. In fact $G(X^b Y^{-a}, X^{-b} Y^A)$ and $F(X, Y)$ become the same polynomial after multiplying by a suitable power of $X$ and $Y$. Write as before $F_\gamma(X, Y)$ for the main part of $F$ corresponding to $\gamma$ and $f_\gamma$, the univariate polynomial corresponding to it. The main part of $G(X, Y)$ corresponding to $\gamma$ is, since we isolated $\gamma$, of the form

$$U^d f_\gamma(U)$$

for some non-negative integer $d$. Remember that since $F$ is non-degenerate with respect to its Newton diagram, the polynomial $f_\gamma$ has only simple roots. Let $\rho$ be such a root. Then we know that

$$G(U, V) = \alpha(U - \rho) + \cdots,$$

where the dots stand for a sum of monomial terms in $U - \rho$ and $V$ which are either divisible by $V$ or of total degree at least 2 and with $\alpha$ a non-zero constant. Hence, we see that $v$ is a uniformizing parameter for the point $P = (\rho, 0)$ on the curve $\mathcal{Y}$. (We write as before small characters for functions and capitals for coordinates.) Note that by definition of $f_\gamma$, we have $f_\gamma(0) \neq 0$. Hence $\rho \neq 0$. This implies that $v_P(u) = 0$. Further we know, since $v$ is a uniformizing parameter for $P$, that $v_P(v) = 1$. This implies $v_P(x) = a$ and $v_P(y) = b$. This concludes the proof. \hfill \Box

**Theorem 3.3.2** Let the curve $\mathcal{X}$ be given by the absolutely irreducible equation $F(X, Y) = 0$ over the perfect field $\mathbb{F}$. Suppose that all singularities of $\mathcal{X}$ are among the points $[0 : 0 : 1]$, $[0 : 1 : 0]$ and $[1 : 0 : 0]$. Further suppose that $F$ is non-degenerate with respect to its Newton polygon. Then the space of regular differential forms has a basis with elements of the form

$$x^iy^j \frac{dx}{xy \frac{dF}{dy}(x, y)},$$

where $i$ and $j$ have the property that $(i, j)$ is an interior integral point of the Newton polygon of $F$.

Before proving this theorem, we state and prove a lemma, which will be needed in the proof. The lemma involves a group action and a certain differential form $\omega_\mathcal{X}$.
3.3 Semigroups of plane curves

**Definition 3.3.3** Let \( \mathcal{X} \) be an algebraic curve over the perfect field \( \mathbb{F} \). Suppose that \( \mathcal{X} \) is a plane curve given by the equation \( F(X, Y) = 0 \), which may be reducible. The basic differential form \( \omega_\mathcal{X} \) corresponding to \( F \) is defined to be

\[
\omega_\mathcal{X} = \frac{dx}{x y \frac{\partial F}{\partial Y}(x, y)}.
\]

When \( \frac{\partial F}{\partial Y}(x, y) \) is identically zero (note that this happens if and only if \( F(X, Y) = G(X, Y^p) \) for some polynomial \( G \)), we define

\[
\omega_\mathcal{X} = -\frac{dy}{x y \frac{\partial F}{\partial X}(x, y)}.
\]

The two definitions coincide if both \( \frac{\partial F}{\partial Y}(x, y) \) and \( \frac{\partial F}{\partial X}(x, y) \) are not zero. To see this, note that since \( F(x, y) = 0 \), we have \( dF(x, y) = 0 \) and hence

\[
\frac{\partial F}{\partial X}(x, y)dx + \frac{\partial F}{\partial Y}(x, y)dy = 0.
\]

The group \( SL(2, \mathbb{Z}) \) of two by two matrices of determinant 1 and coefficients in \( \mathbb{Z} \) acts on the basic differential as we will explain now. Let \( M \) be an element of \( SL(2, \mathbb{Z}) \) and write

\[
M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.
\]

Associated with \( M \), there is the birational map \( \Phi_M \) defined by

\[
\Phi_M(X, Y) = (X^a Y^c, X^b Y^d).
\]

Denote the image under \( \Phi_M \) of \( \mathcal{X} \) by \( \mathcal{Y} \), so we find as map \( \Phi_M : \mathcal{X} \to \mathcal{Y} \). It is easy to check that this is a group action. We can use \( \Phi^{-1}_M : \mathcal{Y} \to \mathcal{X} \) to pull back differential forms on \( \mathcal{X} \) to \( \mathcal{Y} \). We denote this map by \( (\Phi_M)_* : \Omega(\mathcal{X}) \to \Omega(\mathcal{Y}) \). This is a group action as well. The following lemma states a property of this action.

**Lemma 3.3.4** Let \( \mathcal{X} \) be a curve over a perfect field \( \mathbb{F} \) given by the absolutely irreducible equation \( F(X, Y) = 0 \). Let \( M \in SL(2, \mathbb{Z}) \) and denote by \( \mathcal{Y} \) the image of \( \mathcal{X} \) under the map \( \Phi_M \). Denote by \( \omega_\mathcal{X} \) the basic differential form of \( \mathcal{X} \). Define

\[
(u, v) = \Phi_M(x, y).
\]

Then we have

\[
(\Phi_M)_*(\omega_\mathcal{X}) = \frac{du}{uv \frac{\partial F \circ \Phi^{-1}_M}{\partial Y}(u, v)},
\]

or, alternatively,

\[
(\Phi_M)_*(\omega_\mathcal{X}) = -\frac{dv}{uv \frac{\partial F \circ \Phi^{-1}_M}{\partial X}(u, v)}.
\]
Proof: The second identity follows directly from the first. Further remark that it suffices to prove the first identity for a set of generators of \( SL(2, \mathbb{Z}) \). As generators we take

\[
\begin{pmatrix}
1 & -1 \\
0 & 1
\end{pmatrix}
\]

and

\[
\begin{pmatrix}
1 & 0 \\
-1 & 1
\end{pmatrix}.
\]

Indeed, these two elements generate \( SL(2, /\mathbb{Z}) \), since

\[
\begin{pmatrix}
1 & -1 \\
0 & 1
\end{pmatrix}^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}
\]

and

\[
\begin{pmatrix}
1 & -1 \\
0 & 1
\end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.
\]

This shows that we can obtain two matrices which are well-known to generate \( SL(2, /\mathbb{Z}) \) (see for example p.220 Proposition 8.3 of (Knapp 1992)). We start with the first generator. In this case \( \Phi_M(X, Y) = (X, Y/X) \). Define \( U = X \) and \( V = Y/X \). Then we have

\[
\frac{du}{uv} \frac{\partial \Phi \Phi^{-1}}{\partial v}(u, v) = \frac{du}{uv} \frac{\partial \Phi(U, V)}{\partial v}(u, v) = \frac{dx}{y \frac{\partial \Phi}{\partial y}(x, y)/x} = \omega_X.
\]

Here we have identified \( \omega_X \) with \( (\Phi_M)_* \omega_X \).

Now consider the second generator. We have \( \Phi_M(X, Y) = (X/Y, Y) \). Define \( U = X/Y \) and \( V = Y \). If \( \frac{\partial \Phi}{\partial v} \neq 0 \), it is convenient to use the other representation of the basic differential form as in Definition 3.3.3. If this is not the case we proceed as follows. We can assume that \( \frac{\partial \Phi}{\partial v}(x, y) \neq 0 \). We have

\[
\frac{du}{uv} \frac{\partial \Phi \Phi^{-1}}{\partial v}(u, v) = \frac{du}{uv} \frac{\partial \Phi(U, V)}{\partial v}(u, v) = \frac{du}{x(u \frac{\partial \Phi}{\partial X}(x, y) + \frac{\partial \Phi}{\partial V}(x, y))}.
\]

Since \( u = x/y \), we have \( du = dx/y - x dy / y^2 \). However, since \( \frac{\partial \Phi}{\partial v}(x, y) \neq 0 \), we have

\[
dy = -\frac{\partial \Phi}{\partial v}(x, y) dx.
\]

Hence we can express \( du \) as a rational function in \( x \) and \( y \) times \( dx \). Using this we find

\[
\frac{du}{uv} \frac{\partial \Phi \Phi^{-1}}{\partial v}(u, v) = \omega_X.
\]
This concludes the proof of the lemma. □

Now we give the proof of Theorem 3.3.2.

Proof: The differential forms mentioned above are clearly independent, since the monomials $x^iy^j$ with $(i, j)$ and interior integral point of the Newton polygon, are linearly independent. However, the number of differential forms we have found in this way is, according to Baker’s theorem, exactly the genus of the curve. Hence we are left to show that the differential forms mentioned above are regular.

Choose an interior integral point $(i, j)$ of the Newton polygon and define $\omega = x^iy^j\omega_X$. First we will calculate $v_P(\omega)$ for the affine points different from $[0 : 0 : 1]$. Write $P = (\alpha, \beta)$. Since $X$ has no singularities outside $[0 : 0 : 1]$, $[0 : 1 : 0]$ or $[1 : 0 : 0]$, we know that either $\frac{\partial F}{\partial Y}(\alpha, \beta)$ or $\frac{\partial F}{\partial Y}(\alpha, \beta)$ is non-zero. Suppose the first is the case (the other case is similar). Then

$$v_P(\omega) = (i - 1)v_P(x) + (j - 1)v_P(y) + v_P(dy) \geq 0.$$ 

Next, we calculate $v_P(P)$ for points at infinity of the form $[1 : \alpha : 0]$, where $\alpha \neq 0$. In this case we need to rewrite $\omega$. Using the homogeneous notation we have for example

$$\omega = \left(\frac{X}{Z}\right)^{i-1} \left(\frac{Y}{Z}\right)^{j-1} Z^{\deg(F)-1} \frac{\partial F}{\partial Y}(X, Y, Z) d\left(\frac{X}{Z}\right)$$

and hence

$$\omega = -\left(\frac{Z}{X}\right)^{\deg(F)-i-1} \left(\frac{Y}{X}\right)^{j-1} X^{\deg(F)-1} \frac{\partial F}{\partial Y}(X, Y, Z) d\left(\frac{Z}{X}\right).$$

Here $F^*(X, Y, Z)$ denotes the homogenized version of $F(X, Y)$. This is not hard to prove and left to the reader. Using the above identity and similar ones, one sees that $v_P(\omega) \geq 0$ for this $P$ as well.

Now we will calculate $v_P(\omega_X)$ for $P$ infinitely near to $[0 : 0 : 1]$ (the case $P$ infinitely near to $[0 : 1 : 0]$ or $[1 : 0 : 0]$ is similar). Suppose that $P$ corresponds to the edge $\gamma$ of length $l(\gamma)$. And suppose that $\gamma$ lies on the line given by the equation

$$ai + bj = c,$$

where $a$ and $b$ are relatively prime positive integers. Using Lemma 2.5.3 we see that that map $\Phi_M$ isolates $\gamma$, where

$$M = \begin{pmatrix} b & -B \\ -a & A \end{pmatrix},$$

as in the proof of Proposition 3.3.1. Define $(u, v) = (x^by^{-a}, x^{-b}y^a)$. We want to use Lemma 3.3.4 to compute $v_P(\omega_X)$. We know from Proposition 3.3.1 that $v_P(u) = 0$ and $v_P(v) = 1$. Hence $v_P(dv) = 0$. Further note that
F \circ \Phi_{\gamma}^{-1}(U, V) \equiv F_{\gamma}(U^AV^a, U^BV^b) \equiv V^cU^d f_{\gamma}(U) \quad (\text{mod } V^{c+1}).

with \( d \) some non-negative integer. Hence

\[
\frac{\partial F \circ \Phi_{\gamma}^{-1}(U, V)}{\partial U} \equiv V^c \left( U^d \frac{df_{\gamma}(U)}{dU} + dU^{d-1} f_{\gamma}(U) \right) \quad (\text{mod } V^{c+1}).
\]

Since \( F \) is non-degenerate with respect to its Newton polygon, we see that

\[ v_P \left( \left. \frac{df_{\gamma}(T)}{dT} \right|_{T=a} \right) = 0. \]

Taking all elements together, we obtain

\[ v_P(\omega_{\mathcal{X}}) = -(c + 1) \]

and hence

\[ v_P(x^i y^j \omega_{\mathcal{X}}) = ai + bj - c - 1. \]

This differential form is regular at \( P \) if and only if \((i, j)\) lies above the line through the edge \( \gamma \). Taking into account all edges, we see that this differential form is regular at all points if and only if \((i, j)\) is an interior integral point of the Newton polygon. This concludes the proof of the theorem.

Using the above theorem and Proposition 3.2.5 gives us a way of obtaining information about semigroups. The following proposition exploits this.

**Proposition 3.3.5** Let the curve \( \mathcal{X} \) be given by the absolutely irreducible equation \( F(X, Y) = 0 \) over the perfect field \( E \). Suppose that all singularities of \( \mathcal{X} \) are among the points \([0 : 0 : 1], [0 : 1 : 0] \) and \([1 : 0 : 0] \). Further suppose that \( F \) is non-degenerate with respect to its Newton polygon \( \Gamma(F) \). Let \( P \) be a point on the non-singular model of \( \mathcal{X} \) corresponding to the edge \( \gamma \) of the Newton polygon. Suppose that \( \gamma \) lies on the line given by the equation

\[ ai + bj = c, \]

where \( a \) and \( b \) are relatively prime integers. Finally suppose that

\[ \Gamma(F) \subset \{ (i,j) \mid ai + bj \geq c \}. \]

If \((i,j)\) is an interior integral point of the Newton polygon, then

\[ ai + bj - c \]

is a gap at \( P \).
3.3 Semigroups of plane curves

**Proof:** Let \((i, j)\) be an interior integral point of the Newton polygon of \(F\) and denote by \(\omega_\lambda\) the basic differential corresponding to \(F\). From the proof of Theorem 3.3.2 we see that the regular differential form

\[ x^i y^j \omega_\lambda \]

has valuation

\[ ai + bj - c - 1 \]

at \(P\). Using Proposition 3.2.5 one obtains the desired result. \(\square\)

**Remark 3.3.6** In the situation of Proposition 3.3.5 not all gaps will be found in general. However, according to Theorem 3.3.2, we do have a basis for the space of regular differential forms. So, all gaps can be found by applying the following principle. Whenever two elements of the basis have the same valuation at \(P\), replace the second by a suitable linear combination of the two such that the valuation at \(P\) becomes larger. After repeating this a finite number of times, one obtains a basis with the property that no two basis elements have the same valuation at \(P\). Taking the valuation at \(P\) of the basis elements will give all gaps of the semigroup \(W_P\). This algorithm is in essence already contained in Volume V, pp. 88-89 of (Baker 1960).

In the following proposition we combine the above algorithm and (the proof of) Proposition 3.3.5 to obtain more information about gaps.

**Proposition 3.3.7** Let the notation be as in Proposition 3.3.5. Suppose that \((i, j)\) and \((k, l)\) are two interior integral points of the Newton polygon such that

\[ ai + bj = ak + bl. \]

Then there exists an integer \(m\) such that

\[ (k - i, l - j) = m(-b, a). \]

Write \(P\) for the point on the non-singular model corresponding to the root \(\rho\) of the polynomial \(f_\lambda(T)\), then the numbers

\[ ai + bj - c + rv_P \left( \frac{x^b}{y^a} - \rho^{-1} \right), \]

are gaps at \(P\), with \(0 \leq r \leq |m|\).

**Proof:** The proof makes use of the algorithm mentioned in the above remark. First note that

\[ ai + bj = ak + bl \]
implies that

\[ a(i - k) \equiv 0 \pmod{b}. \]

Since \( a \) and \( b \) are relatively prime, this implies

\[ i \equiv k \pmod{b}. \]

Hence, an integer \( m \) exists such that \( k - i + mb = 0 \). Hence, using the equation \( ai + bj = ak + bl \), we find

\[ (k - i, l - j) = m(-b, a). \]

We can assume that \( m > 0 \) by interchanging, if necessary, the roles of \((i, j)\) and \((k, l)\). We see that the points

\[ (i, j), \ldots, (i - rb, j + ra), \ldots, (i - mb, j + ma) \]

are interior integral points of the Newton polygon. Let us apply the algorithm mentioned in the above remark to the corresponding regular differential forms. We are going to replace the differential form

\[ x^{i-b}y^{j+a} \omega_X \]

by

\[ (x^{i-b}y^{j+a} - \rho^{-1}x^{i}y^{j}) \omega_X, \]

where \( \rho^{-1} \) is the root of \( f_\gamma(T) \) corresponding to \( P \). To calculate the valuation at \( P \) of this differential form we use the same birational map as we did in the proof of Theorem 3.3.2. This gives

\[ v_P((x^{i-b}y^{j+a} - \rho^{-1}x^{i}y^{j}) \omega_X) = ai + bj - c - 1 + v_P(u - \rho^{-1}), \]

where \( u = x^b/y^a \). Hence the number

\[ ai + bj - c + v_P(u - \rho^{-1}) \]

is a gap. Repeating the argument, we conclude that the number

\[ ai + bj - c + rv_P(u - \rho) \]

is a gap at \( P \). □
3.4 Semigroups of plane curves of type II

In (Feng and Rao 1994) curves of type I are investigated. These are curves given by the equation

$$Y^b + uX^a + G(X, Y) = 0.$$ 

Here $a$ and $b$ are relative prime numbers and $u$ a non-zero constant from a perfect field $\mathbb{F}$. Further $G$ satisfies $\text{wdeg}(G) < ab$ and $\text{deg}_X(G) < a$. The weighted degree $\text{wdeg}$ is defined by $\text{wdeg}(X) = b$ and $\text{wdeg}(Y) = a$. It is often assumed that the curve has no singularities in its affine part. As a generalization, curves with equation

$$Y^{b+c} + uX^aY^c + G(X, Y) = 0,$$

are considered in Volume I, Chapter 10, pages 907-910 of (Pless, Huffman, and Brualdi 1998). Here $a$, $b$, $u$ and $G$ are as before, while $c$ is a positive number. Usually one assumes that $a > b$ and the semigroup and/or the ring of functions regular outside the point $P = [0 : 1 : 0]$ is investigated. These curves are called curves of type II. Again it is usually assumed that the curve has no singularities in its affine part.

![Figure 3.1: Newton diagram of a type II curve](image-url)
It is true that a curve of type II is birationally equivalent to a curve of type I. However, the corresponding curve of type I will usually have singularities in its affine part, so most of the known theory of type I curves does not apply to those of type II. The Newton polygon of a curve of type II has a shape as depicted in Figure 3.1.

Since the numbers $a$ and $b$ are relatively prime, we see that the edge of the Newton polygon connecting $(0, b + c)$ and $(a, c)$ corresponds to exactly one point on the non-singular model. Denote this point by $P$. If $a > b$, then $P$ corresponds to the point $[0 : 1 : 0]$ of $X$. This explains the fact that usually it is assumed that $a > b$, but we will not make this assumption. Using the results in the Proposition 3.3.5, we obtain the following theorem.

**Theorem 3.4.1** Let $X$ be a curve of type II given by the equation

$$Y^{b+c} + uX^aY^c + G(X,Y) = 0.$$  

As before suppose that $a$ and $b$ are relatively prime positive integers, $c$ is a non-negative integer and $G(X,Y)$ is a bivariate polynomial satisfying $\text{wdeg}(G) < ab$ and $\text{deg}_{X}(G) < a$. Here the weighted degree $\text{wdeg}$ is defined by $\text{wdeg}(X) = b$ and $\text{wdeg}(Y) = a$. Finally suppose that $u$ denotes some non-zero constant from the field $\mathbb{F}$ of definition. As always we assume that the field $\mathbb{F}$ is perfect. Denote by $P$ the point corresponding to the edge of the Newton polygon connecting $(0, b+c)$ and $(a, c)$. Suppose that

$$F(X,Y) = Y^{b+c} + uX^aY^c + G(X,Y)$$

is non-degenerate with respect to its Newton polygon $\Gamma(F)$ and that the singularities of $X$ are among $[0 : 0 : 1]$, $[0 : 1 : 0]$ and $[1 : 0 : 0]$. The gaps of the semigroup $W_P$ are given by

$$-bi - aj + a(b + c),$$

where $(i, j)$ runs over the interior integral points of $\Delta(F)$.

**Proof:** By Proposition 3.3.5 we know that

$$-bi - aj + a(b + c)$$

is a gap if $(i, j)$ is an interior integral point of the Newton polygon. Now we will prove that all these gaps are different. Hence suppose that $(k, l)$ is another interior integral point of $\Gamma(F)$ and that

$$-bi - aj + a(b + c) = -bk - dl + a(b + c).$$

This implies that $b(k - i) + a(l - j) = 0$ and hence

$$b(k - i) \equiv 0 \pmod{a}.$$  

Since $a$ and $b$ are relatively prime this implies
3.4 Semigroups of plane curves of type II

\[ k - i \equiv 0 \pmod a \]

and hence

\[ k = i \text{ or } |k - i| \geq a. \]

On the other hand, since \( \text{deg}_X(G) < a \), we see \( |k - i| < a \). Hence \( k = i \) which implies that \( j = l \) as well. From Baker’s theorem and the Weierstrass gap theorem we conclude that we have found all gaps for the semigroup \( W_P \).

Recall that we mentioned in Corollary 3.2.3 that on a curve of genus \( g \) the largest gap cannot exceed \( 2g - 1 \). Now we have the following statement about the largest gap.

**Corollary 3.4.2** Let the notation and the assumptions be as in the above theorem. Then the largest gap in the semigroup \( W_P \) cannot exceed \( a(b + c) - a - b \).

**Proof:** This follows by remarking that \((1,1)\) can be an interior integral point, but any other integral point \((i, j)\) giving a larger value of the function \(-bi - aj + a(b + c)\) cannot be an interior integral point of the Newton polygon. \( \square \)

It is not always convenient to describe a semigroup by its gaps. Often one wants to know the generators of the semigroup. This makes it easier to check whether or not one has found all functions. It is in principle easy to derive a set of generators from the set of gaps. First, one looks for the smallest non-zero element in the semigroup. This is the first generator. Then one looks inductively for the smallest element in the semigroup which is not in the semigroup generated by the set of generators which have been found before. This process stops after a finite number of steps when the semigroup generated by the set of generators found, equals the whole semigroup. In the following proposition we give a set of generators for the semigroup \( W_P \). First we need a definition.

**Definition 3.4.3** Let \( \Gamma(F) \) be the Newton polygon of a polynomial \( F(X,Y) \) of the form

\[ F(X,Y) = Y^{b+c} + uX^aY^c + G(X,Y), \]

where \( a, b, c \) and \( G \) are as before. Define for \( 0 \leq i < a \) the function \( m_F \) by

\[ m_F(i) = \max\{j \mid j \not\in \Gamma(F)^\circ, bi + aj \leq a(b + c)\}. \]

Here \( \circ \) means taking the interior. In words this means that given an \( i \), the function \( m_F \) gives the largest \( j \) such that \((i, j)\) lies below or on the boundary of the Newton polygon (recall that we assume \( \gcd(a,b) = 1 \)).

**Proposition 3.4.4** Let \( \mathcal{X} \) be a curve over a perfect field \( \mathbb{F} \) of type II given by the equation
\[ Y^{b+c} + uX^aY^c + G(X, Y) = 0. \]

As before suppose that \( a \) and \( b \) are relatively prime positive integers, \( c \) is a non-negative integer and \( G(X, Y) \) is a bivariate polynomial satisfying \( \text{wdeg}(G) < ab \) and \( \text{deg}_X(G) < a \). Here the weighted degree \( \text{wdeg} \) is defined by \( \text{wdeg}(X) = b \) and \( \text{wdeg}(Y) = a \). Finally, suppose that \( u \) denotes some non-zero constant from the field \( F \) of definition. Denote by \( P \) the point corresponding to the edge of the Newton polygon connecting \((0, b + c)\) and \((a, c)\). Suppose that

\[ F(X, Y) = Y^{b+c} + uX^aY^c + G(X, Y) \]

is non-degenerate with respect to its Newton polygon \( \Gamma(F) \) and that the singularities of \( \mathcal{X} \) are among \( [0 : 0 : 1] \), \( [0 : 1 : 0] \) and \( [1 : 0 : 0] \).

The semigroup \( W_P \) is generated by the elements

\[ a \text{ and } a(b + c) - bi - am_F(i), \]

with \( 0 < i < a \).

**Proof:** First note that by Proposition 3.3.5 we know all gaps and that the values \( a \) and \( a(b + c) - bi - am_F(i) \) are not among them for any \( 0 < i < a \), because \((i, m_F(i))\) is not an interior integral point of \( \Delta(F) \). It is not hard to see that the set

\[ \{a(b + c) - bi - aj \mid 0 \leq i < a, bi + aj \leq a(b + c)\} \]

is equal to the set of non-negative integers. In this way, we obtain a bijection between the set of non-negative integers and the set \( S \) of pairs \((a - i, c - j)\) with \( 0 \leq i < a \) and \( bi + aj \leq a(b + c) \). In fact, \( S \) is a semigroup if we define the addition by adding pairs modulo \((a, -b)\). In other words the isomorphism of groups

\[ \mathbb{Z} \cong \mathbb{Z}^2/\langle (a, -b) \rangle \]

restricts to the isomorphism of semigroups

\[ \mathbb{N}_0 \cong S. \]

Moreover, the gaps of \( W_P \) correspond to the pairs \((a - i, c - j)\), where \((i, j)\) is an interior integral point of \( \Delta(F) \). Hence we define \( S' \subset S \) by

\[ S' = \{(a - i, c - j) \in S \mid (i, j) \notin \Delta(F)^o\}. \]

The isomorphism of semigroups \( \mathbb{N}_0 \cong S \) restricts to

\[ W_P \cong S'. \]

From the above it is clear that to prove the proposition, we only need to prove that the pairs \((0, 1)\) and \((a - i, c - m_F(i))\) with \( 0 < i < a \), generate \( S' \) (when adding mod \((a, -b)\)).

Note that for each \( 0 < i < a \), the value \( c - m_F(i) \) is the smallest value \( k \) such that \((a - i, k) \in S' \). Combining these generators with \((0, 1)\) gives all of \( S' \). □
Note that the set of generators in the above proposition need not be minimal. This means that in a specific case it can happen that some of the generators mentioned above can be left out.

3.5 Application to the first Garcia-Stichtenoth tower

In the previous two sections we have investigated semigroups of plane curves. Now we look at the semigroups of the curves arising in the first Garcia-Stichtenoth tower. These curves are examples of affine complete intersections. This means that an affine piece of the curve can be described as the intersection of $n-1$ hypersurfaces. In other words: an affine piece of the curve can be given by $n-1$ (affine) equations. We will follow a similar path as we did in the previous sections. This means that we first find regular differential forms as we did in Theorem 3.3.2 and then apply Proposition 3.2.5 to obtain information about the gaps of a semigroup. Again we try the differential forms mentioned in (Hovanskii 1978) and prove that they are regular in this case as well. We will first define an analog of the basic differential form $\omega_X$ as in Definition 3.3.3 and prove analogous properties about it. Since a lot of the theory is applicable to any complete intersection, we will present the theory more generally at first and then specialize to the first Garcia-Stichtenoth tower. First we introduce some notation.

Definition 3.5.1 Let $F = (F_1, \ldots, F_{n-1})$ be a vector of $n-1$ polynomials in $n$ variables $X_1, \ldots, X_n$. We define

$$J_i(F) = (-1)^{n-i} \begin{vmatrix} \frac{\partial F_1}{\partial X_1} & \cdots & \frac{\partial F_1}{\partial X_{n-1}} & \frac{\partial F_1}{\partial X_n} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial F_{n-1}}{\partial X_1} & \cdots & \frac{\partial F_{n-1}}{\partial X_{n-1}} & \frac{\partial F_{n-1}}{\partial X_n} \end{vmatrix}.$$ 

Now we define the analogue of the basic differential form.

Definition 3.5.2 Let $X$ be an absolutely irreducible algebraic curve over a perfect field $F$ defined by the equation

$$F(X_1, \ldots, X_n) = 0,$$

where $F$ is a vector of polynomials $F_1, \ldots, F_{n-1}$ in $n$ variables. The basic differential form of $X$ is defined to be

$$\frac{dx_i}{x_1 \cdots x_n J_i(F)(x_1, \ldots, x_n)},$$

where $i$ is such that $J_i(F) \neq 0$ as an element of the function field of $X$.

Of course we have to show that such an $i$ exists and that the choice of $i$ does not matter. For the existence note that if for all $1 \leq i \leq n$ the polynomial $J_i(F)$ is the
zero of the function field, then all points of \( \mathcal{X} \) are singular points. This is impossible, since the number of singularities of a curve is always finite. Now suppose that \( i \) and \( j \) are such that both \( J_i(F) \) and \( J_j(F) \) are not the zero of the function field. Since \( F(x_1, \ldots, x_n) = 0 \), we have

\[
\begin{pmatrix}
\frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial F_n}{\partial x_1} & \cdots & \frac{\partial F_n}{\partial x_n}
\end{pmatrix}
\begin{pmatrix}
dx_1 \\
\vdots \\
dx_n
\end{pmatrix}
= 0.
\]

This can be rewritten as

\[
\begin{pmatrix}
\frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_{i-1}} & \frac{\partial F_1}{\partial x_{i+1}} & \cdots & \frac{\partial F_1}{\partial x_n} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\frac{\partial F_n}{\partial x_1} & \cdots & \frac{\partial F_n}{\partial x_{i-1}} & \frac{\partial F_n}{\partial x_{i+1}} & \cdots & \frac{\partial F_n}{\partial x_n}
\end{pmatrix}
\begin{pmatrix}
dx_1 \\
\vdots \\
dx_i \\
\vdots \\
dx_n
\end{pmatrix}
= -\begin{pmatrix}
\frac{\partial F_i}{\partial x_1} \\
\vdots \\
\frac{\partial F_i}{\partial x_n}
\end{pmatrix}
dx_i.
\]

Using Cramer’s rule we find

\[
J_i(F)(x_1, \ldots, x_n)dx_j = J_j(F)(x_1, \ldots, x_n)dx_i,
\]

which is exactly what we need.

Now we will prove a lemma similar to Lemma 3.3.4. Note that \( SL(n, \mathbb{Z}) \) acts on differential forms in a similar way as for the case \( n = 2 \). As before we denote by \( \Phi_M \) the birational map associated with a matrix \( M \in SL(n, \mathbb{Z}) \) and by \( (\Phi_M)^* \) the induced action on differential forms. Further define \( (u_1, \ldots, u_n) = \Phi_M(x_1, \ldots, x_n) \).

**Lemma 3.5.3** Let \( \mathcal{X} \) be an absolutely irreducible curve defined over a perfect field \( F \) whose affine part is given by the vector equation \( F(X_1, \ldots, X_n) = 0 \). Let \( M \in SL(n, \mathbb{Z}) \) and define \( \mathcal{Y} \) to be the image of \( \mathcal{X} \) under the map \( \Phi_M \). Denote by \( \omega_\mathcal{X} \) the basic differential form on \( \mathcal{X} \). We have

\[
(\Phi_M)^*(\omega_\mathcal{X}) = \frac{du_n}{u_1 \cdots u_n J_n(F \circ \Phi_M^{-1})(u_1, \ldots, u_n)}.
\]

**Proof:** It suffices to prove this identity for \( M \) in a set of generators of \( SL(n, \mathbb{Z}) \). Denote by \( E_{i,j} \) the \( n \times n \) matrix all of whose entries are zero except the one in the \( i \)-th column and the \( j \)-th row, which is one instead. The group \( SL(n, \mathbb{Z}) \) is generated by elements of the form \( I + E_{i,j} \), where \( I \) denotes the \( n \times n \) identity matrix and \( i \neq j \) (see for example p.541, Proposition 9.1 of (Lang 1993)). Further note that \( (I + E_{i,j})^{-1} = I - E_{i,j} \). We assume that \( J_n(F) \neq 0 \), but the proof can easily be modified if this is not the case. Write \( M = I + E_{i,j} \) for some \( i \) and \( j \) not equal to each other.
We find the following identities on differential operators

\[
\frac{\partial G(X_1, \ldots, X_n)}{\partial X_k} = \begin{cases} 
\frac{\partial G \circ \Phi^{-1}_M(U_1, \ldots, U_n)}{\partial U_k} & \text{if } k \not\in \{i, j\}, \\
U_j \frac{\partial G \circ \Phi^{-1}_M(U_1, \ldots, U_n)}{\partial U_i} & \text{if } k = i, \\
\frac{U_i \partial G \circ \Phi^{-1}_M(U_1, \ldots, U_n)}{U_j} + \frac{\partial G \circ \Phi^{-1}_M(U_1, \ldots, U_n)}{\partial U_j} & \text{if } k = j
\end{cases}
\]

Further, we have

\[x_1 \cdots x_n = \frac{u_1 \cdots u_n}{u_j}.
\]

First we suppose that \(i \neq n\). In this case \(u_n = x_n\) and hence \(du_n = dx_n\). Here we have identified the function field of \(Y\) with a subfield of \(X\). Using the above identities on differential operators, we have

\[J_n(F)(x_1, \ldots, x_n) = u_j J_n(F \circ \Phi^{-1}_M)(u_1, \ldots, u_n).
\]

Putting everything together we find the desired result.

If \(i = n\), the situation is slightly more complicated. We have \(x_n = u_n/u_j\) and hence

\[dx_n = \frac{du_n}{u_j} - \frac{u_n du_j}{u_j^2}.
\]

Using Cramer’s rule in a similar way as we did after Definition 3.5.2, we find

\[du_j = \frac{J_n(F \circ \Phi^{-1}_M)(u_1, \ldots, u_n)}{J_n(F \circ \Phi^{-1}_M)(u_1, \ldots, u_n)} du_n \]

and hence

\[dx_n = \left(1 - \frac{u_n J_n(F \circ \Phi^{-1}_M)(u_1, \ldots, u_n)}{u_j^2 J_n(F \circ \Phi^{-1}_M)(u_1, \ldots, u_n)}\right) du_n.
\]

Further

\[J_n(F)(x_1, \ldots, x_n) = J_n(F \circ \Phi^{-1}_M)(u_1, \ldots, u_n) - \frac{u_n}{u_j} J_n(F \circ \Phi^{-1}_M)(u_1, \ldots, u_n).
\]

Putting everything together, we obtain the desired result in this case as well.

As we did before we can now get information about semigroups by trying to find regular differential forms of the form \(x_1^{i_1} \cdots x_n^{i_n} \omega_X\). We specialize to the following case. Let \(F(X, Y)\) be an absolutely irreducible polynomial in two variables over a (perfect) field \(F\). Define

\[F_i(X_1, \ldots, X_n) = F(X_i, X_{i+1}).
\]
Define the curve $\mathcal{X}_n$ to be the curve given by equations

$$F_i(X_1, \ldots, X_n) = 0,$$

where $1 \leq i < n$. If we look at the function fields of these curves, we find a tower of function fields $\mathcal{F}_n$ defined by $\mathcal{F}_1 = \mathcal{F}(x_1)$ and $\mathcal{F}_{k+1} = \mathcal{F}_k(x_{k+1})$, where $F(x_k, x_{k+1}) = 0$. We calculate for this case the $J_i$ in the following lemma.

**Lemma 3.5.4** Let $F(X, Y)$ be a bivariate polynomial and define

$$F_i(X_1, \ldots, X_n) = F(X_i, X_{i+1}),$$

where $1 \leq i \leq n - 1$. Using notation as in Definition 3.5.1, we have

$$J_k(F)(X_1, \ldots, X_n) = (-1)^{n-k} \prod_{j=1}^{k-1} \frac{\partial F}{\partial X}(X_j, X_{j+1}) \prod_{j=k}^{n-1} \frac{\partial F}{\partial Y}(X_j, X_{j+1}),$$

with $1 \leq k \leq n$.

**Proof:** This is clear from the definition. \qed

Now we turn to the first Garcia-Stichtenoth tower. In (García and Stichtenoth 1995) the following tower is considered. Let $F = F_{q^2}$ and

$$F(X, Y) = X^{q-1}Y^q + Y - X^q.$$

The tower one obtains with this $F$ is asymptotically good. The genus $g_n$ of $\mathcal{X}_n$ satisfies

$$g_n = \begin{cases} 
q^n + q^{n-1} - q^{\frac{n+1}{2}} - 2q^{\frac{n-1}{2}} + 1 & \text{if } n \equiv 1 \pmod{2} \\
q^n + q^{n-1} - \frac{1}{2}q^{\frac{n+1}{2}} - \frac{3}{2}q^{\frac{n-1}{2}} + 1 & \text{if } n \equiv 0 \pmod{2}
\end{cases}$$

For the number of $F_{q^2}$-rational points $N_1(\mathcal{X}_n)$ one has for $n \geq 3$

$$N_1(\mathcal{X}_n) \geq (q^2 - 1)q^{n-1} + 2q.$$

One sees that

$$\lim_{n \to \infty} \frac{N_1(\mathcal{X}_n)}{g_n} = q - 1.$$

This was the first explicitly known tower attaining the Drinfeld-Vlăduţ bound as was mentioned in Chapter 1. Now we return to the investigation of the basic differential form for curves in this tower. We see from the previous lemma that for this tower
3.5 Application to the first Garcia-Stichtenoth tower

\[ J_k(F) = (-1)^{n-k} \prod_{j=1}^{k-1} (-X_j^{q^k-2} X_j^q). \]

We will need some facts mentioned in (García and Stichtenoth 1995). We have defined the tower in terms of \( x_i \). In (García and Stichtenoth 1995) the towers are defined in terms of \( z_i \), which can be defined in terms of \( x_i \) by

\[ z_{i+1} = x_i x_{i+1} \]

for all \( i > 0 \). Note that from

\[ x_{i+1}^{q_i} x_i^{q_i-1} + x_i = x_i^{q_i} \]

we can deduce

\[ z_{i+1}^{q_i} + z_i^{q_i-1} = x_i^{q_i}, \]

which explains the importance of the \( z_i \), for it is now apparent that the tower is formed by taking successive Artin-Schreier extensions. Denote the \( n \)-th function field in the tower by \( \mathcal{F}_n \). Denote the set of places corresponding to this function field by \( \mathcal{P}(\mathcal{F}_n) \). We cite Lemma 2.3 of (García and Stichtenoth 1995).

**Lemma 3.5.5** For all \( n \geq 1 \) there exists a unique place \( Q_n \in \mathcal{P}(\mathcal{F}_n) \) which is a common zero of the functions \( x_1, z_2, z_3, \ldots, z_n \). Its degree is one and we have

\[ v_{Q_n}(x_j) = q^{i-1}. \]

In the extension \( \mathcal{F}_{n+1}/\mathcal{F}_n \) the place \( Q_n \) splits into \( q \) places of degree one (one of them being \( Q_{n+1} \)).

The next lemma is also contained (although not literally) in (García and Stichtenoth 1995).

**Lemma 3.5.6** Denote by \( P_1 \) the place in \( \mathcal{P}(\mathcal{F}_1) \) which is a pole of the function \( x_1 \). There exists a unique point \( P_n \) in \( \mathcal{P}(\mathcal{F}_n) \) lying above \( P_1 \). Its degree is one. In the extension \( \mathcal{F}_{n+1}/\mathcal{F}_n \) the place \( P_n \) is totally ramified. Further we have

\[ v_{P_n}(x_n) = -1. \]

If \( P \in \mathcal{P}(\mathcal{F}_{n+1}) \), denote by \( P \cap \mathcal{F}_n \) the restriction of \( P \) to \( \mathcal{F}_n \). Now we define some sets of places.

**Definition 3.5.7** For \( n \geq 2 \) let

\[ S_0^{(n)} = \{ P \in \mathcal{P}(\mathcal{F}_n) \mid P \cap \mathcal{F}_{n-1} = Q_{n-1} \text{ and } P \neq Q_n \}. \]

Further for \( 1 \leq i \leq n-2 \) define

\[ S_i^{(n)} = \{ P \in \mathcal{P}(\mathcal{F}_n) \mid P \cap \mathcal{F}_{n-1} \in S_{i-1}^{(n-1)} \}. \]
We cite some more facts in the following (see Proposition 2.5, Lemma 2.8 and the remarks following Lemma 2.9 in (Garcia and Stichtenoth 1995)).

**Proposition 3.5.8** Let $P$ be a place in $S_i^{(n)}$.
If $0 \leq i \leq \lfloor \frac{n-2}{2} \rfloor$ then the place $P$ is unramified in the extension $\mathcal{F}_{n+1}/\mathcal{F}_n$ and

$$v_P(x_n) = -q^{n-2i-2}.$$ 

If $\lfloor \frac{n-3}{2} \rfloor < i \leq n-2$, then $P$ is totally ramified in the extension $\mathcal{F}_{n+1}/\mathcal{F}_n$ and

$$v_P(x_n) = -1.$$ 

Further, if $P \in \mathcal{P}(\mathcal{F}_n)$ ramifies in the extension $\mathcal{F}_{n+1}/\mathcal{F}_n$ then either $P = P_n$ or $P \in S_i^{(n)}$ with $\lfloor \frac{n-3}{2} \rfloor < i \leq n-2$.

Figure 3.5 gives the relation between the places mentioned above. The thin lines indicate unramified behaviour, while the thick lines stand for totally ramified behaviour.

![Diagram](image)

The following proposition gives an expression for the valuation of $x_j$ in several places.

**Proposition 3.5.9** We have

$$v_{Q_n}(x_j) = q^{j-1}$$
3.5 Application to the first Garcia-Stichtenoth tower

and

\[ v_{P_n}(x_j) = -q^{n-j}. \]

Let \( P \in S_i^{(n)} \) be a place. Then if \( 0 \leq i \leq \left\lfloor \frac{n-3}{2} \right\rfloor \) we have

\[ v_P(x_j) = \begin{cases} 
-q^{2(n-i)-j-2} & \text{if } j \geq n-i, \\
q^{j-1} & \text{if } j < n-i, 
\end{cases} \]

while for \( \left\lfloor \frac{n-3}{2} \right\rfloor < i \leq n-2 \) we have

\[ v_P(x_j) = \begin{cases} 
-q^{n-j} & \text{if } j \geq n-i, \\
q^{2n-n+j+1} & \text{if } j < n-i. 
\end{cases} \]

**Proof:** This is clear from the above lemmas. The statement about \( Q_n \) is proved in Lemma 3.5.5. The valuation of \( x_n \) in all other relevant points is proved in Lemma 3.5.6 and Proposition 3.5.8. Since we know the ramification structure (as mentioned in Proposition 3.5.8), it is not hard to calculate the valuation of the other \( x_j \) by induction. \( \square \)

**Remark 3.5.10** From the above proposition we can see that there exists symmetry, namely

\[ v_{Q_n}(x_j) = -v_{P_n}(x_{n-j}). \]

Further if we choose \( P \in S_i^{(n)} \) and \( Q \in S_{n-i-2} \), we have

\[ v_P(x_j) = -v_Q(x_{n-j}). \]

The reason for this is that there exists an involution \( \iota_n \) on the function field \( \mathcal{F}_n \) defined by

\[ \iota_n(x_j) = \frac{1}{x_{n-j}}, \]

for \( 1 \leq j \leq n \). Indeed from the known equation

\[ x_i^{q-1}x_{i+1}^q + x_{i+1} = x_i^q, \]

we find after dividing by \( x_i^q x_{i+1}^q \) the equation

\[ \left( \frac{1}{x_{i+1}} \right)^{q-1} \left( \frac{1}{x_i} \right)^q + \frac{1}{x_i} = \left( \frac{1}{x_{i+1}} \right)^q, \]

which is of the same form. This involution interchanges the points \( P_n \) and \( Q_n \) and the sets \( S_i^{(n)} \) and \( S_{n-i-2}^{(n)} \).

A direct consequence of the above remark is the following lemma.
Lemma 3.5.11 Denote by $Q_n$ (respectively $P_n$) the points defined in Lemma 3.5.5 (respectively Lemma 3.5.6). The semigroups of the points $P_n$ and $Q_n$ are the same.

Now we return to our original point of view, namely viewing the curve $X_n$ as (the projective closure of) a curve embedded in an $n$-dimensional affine space given by $n - 1$ equations. The function field of $X_n$ is $\mathcal{F}_n$. The curve $X_n$ has singularities and our first goal is to track these down. We will do this in the next lemma and give some other properties as well.

Lemma 3.5.12 Suppose $n \geq 2$. The set of singularities of the (projective closure of the) curve $X_n$ is given by

$$\{(1:0: \cdots :0), (0:1:0: \cdots :0), \ldots, (0: \cdots :0:1:0)\}.$$ 

If $(\alpha_1, \ldots, \alpha_n)$ is an affine point of $X_n$, then either

$$(\alpha_1, \ldots, \alpha_n) = (0, \ldots, 0),$$

or

$$\prod_{i=1}^{n} \alpha_i \neq 0.$$ 

Proof: We know that an affine part of $X_n$ is given by the equations

$$X_{i+1}^q - X_i^{q+1} + X_{i+1} = X_i^q,$$

where $1 \leq i \leq n - 1$. First we investigate the affine points. Let $(\alpha_1, \ldots, \alpha_n)$ be such a point. If $\alpha_i = 0$, we see immediately from the equations that the other coordinates are zero as well. This proves the second part of the lemma. Further, since $J_1(X_1, \ldots, X_n) = (-1)^{n-1}$ (by Lemma 3.5.4) no affine point can be singular.

Now we investigate the points at infinity of $X_n$. Note that it is not sufficient to homogenize the $n - 1$ affine equations and determine their solutions at infinity, since (the projective closure of) $X_n$ is not a projective complete intersection. So instead, we use the description as given in (García and Stichtenoth 1995) of the places of the function field $\mathcal{F}_n$. These places correspond to the points of the non-singular model of $X_n$. To find the projective coordinates of such a place $P$ in the model $X_n$, it suffices to evaluate the coordinate functions $x_1$ up to $x_n$ at $P$. When one of the $x_j$ has a pole at $P$, one has to multiply the ratio of functions $[x_1: \cdots :x_n:1]$ by a function $f$ such that no $x_if$ has a pole at $P$, but not all the $x_if$'s evaluate to zero either. This is always possible. The projective coordinates in the model $X_n$ corresponding to the place $P$ are then given by

$$[(x_1f)(P) : \cdots : (x_nf)(P) : f(P)],$$

Using this, it is not hard to see that the place $Q_n$ corresponds to the point $[0 : \cdots : 0 : 1]$ and $P_n$ to the point $[1 : 0 : \cdots : 0]$. We simply use Proposition
3.5 Application to the first Garcia-Stichtenoth tower

3.5.9. In a similar way one sees that a place from the set $S_i^{(n)}$ corresponds to the point $[0 : \cdots : 0 : 1 : 0 : \cdots : 0]$, where the 1 is on the $(n-i)$-th position. Since the cardinality of the sets $S_i^{(n)}$ is larger than one, this means that several places will correspond to the same point of $\mathcal{X}_n$. This implies that the points $[0 : 1 : 0 : \cdots : 0]$, $[0 : \cdots : 0 : 1 : 0]$ are singular points of $\mathcal{X}$. Since all other places of $\mathcal{F}_n$ correspond to affine points, we have determined all points at infinity of $\mathcal{X}_n$. What remains to be proved is that $P_n = [1 : 0 : \cdots : 0]$ is a singular point. This can be done by looking at the coordinate functions of $P_n$. For an affine point the coordinate functions are given by $x_j$, where $1 \leq j \leq n$. Hence for $P_n$ they are given by $1/x_1$ and $x_j/x_1$, where $2 \leq j \leq n$. If a point is regular, the valuation of one of the coordinate functions will be one. However, using Proposition 3.5.9, we see that this is not the case for $P_n$.

We now return to the investigation of the differential forms of type $x_1^{i_1} \cdots x_n^{i_n} \omega_{\mathcal{X}_n}$, where $\mathcal{X}_n$ still denotes the $n$-th curve from the first Garcia-Stichtenoth tower.

**Proposition 3.5.13** The differential form $\omega = x_1^{i_1} \cdots x_n^{i_n} \omega_{\mathcal{X}_n}$ is regular outside the $n + 1$ points

$[1 : 0 : \cdots : 0], [0 : 1 : 0 : \cdots : 0], \ldots, [0 : \cdots : 0 : 1]$.

We have

$$v_{Q_n}(\omega) = \sum_{j=1}^{n} (i_j - 1)q^{j-1},$$

while for any $P \in S_i^{(n)}$ for $0 \leq \lfloor (n-3)/2 \rfloor$ we have

$$v_P(\omega) = \sum_{j=1}^{n-i-1} (i_j - 1)q^{j-1} - \sum_{j=n-i}^{n} (i_j - 1)q^{2(n-i)-j-2}.$$  

Further

$$v_{P_n}(\omega) = -(i_1 - q + 1)q^{n-1} - \sum_{j=2}^{n-1} (i_j - 2q + 1)q^{n-j} -(i_n - q + 1)$$

and for any $P \in S_i^{(n)}$ with $\lfloor (n-3)/2 \rfloor < i \leq n - 2$

$$v_P(\omega) = (i_1 - q + 1)q^{2i-n+2} + \sum_{j=2}^{n-i-1} (i_j - 2q + 1)q^{2i-n+j+1} - \sum_{j=n-i}^{n-1} (i_j - 2q + 1)q^{n-j} -(i_n - q + 1).$$

**Proof:** We know by Lemma 3.5.4 that

$$\omega_{\mathcal{X}_n} = \frac{(-1)^{n-i} dx_1}{x_1 \cdots x_n}.$$
Using Lemma 3.5.12 we see that this differential form is regular at any affine point of $\mathcal{X}_n$ not equal to $(0, \ldots, 0)$. This proves the first part of the lemma. For the second part we use Proposition 3.5.9. Using the above form of $\omega_{\mathcal{X}_n}$ we find the values mentioned above of $v_P(\omega_{\mathcal{X}_n})$ for $P = Q_n$ or $P \in S_1^{(n)}$ if $0 \leq i \leq [(n - 3)/2]$. Using the equation
\[
\omega_{\mathcal{X}_n} = \frac{(-1)^{n-1} dx_n}{x_1^{q-1} x_2^{q-1} \cdots x_n^{q-1} x_n},
\]
we find the rest of the proposition.

\[\square\]

**Corollary 3.5.14** The differential form $x_1^{i_1} \cdots x_n^{i_n} \omega_{\mathcal{X}_n}$ is regular on $\mathcal{X}_n$ if and only if the following inequalities hold.

\[
\sum_{j=1}^{n} (i_j - 1)q^{j-1} \geq 0,
\]

\[
\sum_{j=1}^{n-i-1} (i_j - 1)q^{j-1} - \sum_{j=n-i}^{n} (i_j - 1)q^{2(n-0)-j-2} \geq 0
\]

for all $0 \leq i \leq [(n - 3)/2],$

\[
(i_1 - q + 1)q^{2i_1-n+2} + \sum_{j=2}^{n-i-1} (i_j - 2q + 1)q^{2i-j+1} - \sum_{j=n-i}^{n-1} (i_j - 2q + 1)q^{n-j} - (i_n - q + 1) \geq 0
\]

for all $[(n - 3)/2] < i \leq n - 2$ and

\[-(i_1 - q + 1)q^{n-1} - \sum_{j=2}^{n-1} (i_j - 2q + 1)q^{n-j} - (i_n - q + 1) \geq 0.
\]

*Further, if an n-tuple satisfies all these inequalities, then the number

\[(q - 1 - i_1)q^{n-1} + \sum_{j=2}^{n-1} (2q - 1 - i_j)q^{n-j} + (q - i_n)\]

is a gap at $P_n$.*

**Proof:** This is immediate from the above proposition and Proposition 3.2.5. \[\square\]

This corollary gives an algorithm to find gaps of the semigroup $W_{P_n}$. Unfortunately this algorithm does not find all gaps in general. We already know that we will find all gaps for $n = 2$ from Theorem 3.4.1. However for $n = 3$, we find all regular differential forms and hence all gaps as well. We will prove this in the next theorem.
3.5 Application to the first Garcia-Stichtenoth tower 75

**Theorem 3.5.15** The space of regular differential forms over the curve $X_3$ is generated by differential forms of the form

$$x_1^{i_1}x_2^{i_2}x_3^{i_3} \omega x_3,$$

where the 3-tuple $(i_1, i_2, i_3)$ satisfies the inequalities

$$(i_1 - 1) + (i_2 - 1)q + (i_3 - 1)q^2 \geq 0,$$

$$(i_1 - 1) + (i_2 - 1)q - (i_3 - 1)q \geq 0,$$

$$(i_1 - q + 1)q - (i_2 - 2q + 1)q - (i_3 - q + 1) \geq 0,$$

and

$$-(i_1 - q + 1)q^2 - (i_2 - 2q + 1)q - (i_3 - q + 1) \geq 0.$$

**Proof:** We know the the curve $X_3$ has genus $q^3 - 2q + 1$, so all we have to do is find $q^3 - 2q + 1$ linearly independent differential forms of the form mentioned above. We will do this by first giving triples $(i_1, i_2, i_3)$ satisfying the inequalities mentioned above. After that we will determine the gaps arising from these triples in the semigroup $W_{Q_3}$. Finally, we will show that we have found all $q^3 - 2q + 1$ gaps. This will be sufficient to prove the theorem.

We will determine a number of solutions $(i_1, i_2, i_3)$ with the following method; first we fix the third coordinate $i_3$ and then we determine all possible $i_1$ and $i_2$. Suppose that $1 \leq i_3 < q$. Given $i_3$, the conditions on $i_1$ and $i_2$ are

$$(i_1 - 1) + (i_2 - 1)q \geq (i_3 - 1)q,$$

$$(i_1 - q + 1)q - (i_2 - 2q + 1)q \geq (i_3 - q + 1),$$

and

$$-(i_1 - q + 1)q^2 - (i_2 - 2q + 1)q - (i_3 - q + 1) \geq 0.$$ 

Hence for a given $i_3$ with $1 \leq i_3 < q$ the integral solutions for $(i_1, i_2)$ are described by

$$i_2 = i_3 \text{ and } 1 \leq i_1 \leq q,$$

or

$$i_1 = q \text{ and } i_3 \leq i_2 \leq q - 1,$$

or

$$i_1 \leq q - 1, \ i_2 \geq i_3 + 1 \text{ and } i_2 - i_1 \leq q.$$
Not all solutions give rise to different gaps, but if we leave out the solutions with \( i_1 \leq 0 \) or \( i_2 > q + i_3 \), this problem is solved. In this way, we find for a fixed \( i_3 \) with \( 1 \leq i_3 < q \), a total of

\[
q^2 + q - 1 - \sum_{j=0}^{i_3} j
\]

different gaps and hence for \( 1 \leq i_3 < q \) in total

\[
(q - 1)(q^2 + q - 1) - \sum_{i_3=1}^{q-1} \sum_{j=0}^{i_3} j
\]

different gaps. We also find the same number of linearly independent regular differential forms.

Now we consider solutions with \( q \leq i_3 < 2q \). In this case, given \( i_3 \), the integral solutions \( (i_1, i_2) \) are described by

\[
i_1 \leq q - 1 \text{ and } i_2 \geq i_3 \text{ and } i_2 - i_1 \leq q - 1.
\]

This gives for a fixed \( i_3 \)

\[
\sum_{j=0}^{2q^2 - 2i_3} j
\]

new gaps. Hence in total for \( q \leq i_3 < 2q \) we find

\[
\sum_{i_3=q}^{2q-1} \sum_{j=0}^{2q^2 - 2i_3} j
\]

new gaps. Adding up the number of gaps corresponding to \( 1 \leq i_3 < q \) and \( q \leq i_3 < 2q \) gives a total of \( q^3 - 2q + 1 \) gaps, which is the genus of the curve. Hence we have found all gaps.

From the above proof it has become clear what the gaps of the semigroup \( W_{Q_3} \) are. We state this as a corollary.

**Corollary 3.5.16** Denote by \( S \) the set of triples \( (i_1, i_2, i_3) \in \mathbb{Z}^3 \) satisfying

\[
1 \leq i_1 \leq q, \ i_3 \leq i_2 \leq q - 1 \text{ and } 1 \leq i_3 < q,
\]
or

\[
1 \leq i_1 \leq q - 1, \ i_2 = q \text{ and } 1 \leq i_3 < q,
\]
or

\[
1 \leq i_1 \leq q - 1, \ q + 1 \leq i_2 \leq q + i_3, \ i_2 - i_1 \leq q \text{ and } 1 \leq i_3 < q,
\]
or
\[ i_1 \leq q - 1, \ i_2 \geq i_3, \ i_2 - i_1 \leq q - 1 \text{ and } q \leq i_3 < 2q - 1. \]

The set of gaps of the semigroup \( W_{Q_3} \) (or \( W_{P_3} \)) is in one to one correspondence with the set \( S \). The map \( \gamma \) defined by
\[ \gamma(i_1, i_2, i_3) = i_1 + q(i_2 - 1) + q^2(i_3 - 1) \]
gives a bijection between \( S \) and the set of gaps.

**Proof:** This is clear from the proof of the last theorem. Note that the semigroups of \( Q_3 \) and \( P_3 \) are the same, as remarked before. \( \square \)

**Corollary 3.5.17** The largest gap of the semigroup \( W_{Q_3} \) (or \( W_{P_3} \)) equals
\[ 2q^3 - q^2 - 2q - 1. \]

**Proof:** The triple \( (q - 1, 2q - 2, 2q - 2) \) corresponds to this gap and any other triple in \( S \) corresponds to a smaller gap. \( \square \)

**Example 3.5.18** Suppose \( q = 2 \). We have
\[ S = \{(1, 1, 1), (1, 2, 1), (2, 1, 1), (1, 2, 1), (1, 3, 1), (2, 1, 2)\} \]
and corresponding set of gaps
\[ \{2, 3, 5, 7\}. \]

Now suppose that \( q = 3 \). In this case we have
\[ S = \{(1, 1, 1), (2, 1, 1), (3, 1, 1), (1, 2, 1), (2, 2, 1), (3, 2, 1), (1, 3, 1), (2, 3, 1), (1, 4, 1), (2, 4, 1), (1, 2, 2), (2, 2, 2), (3, 2, 2), (1, 3, 2), (2, 3, 2), (1, 4, 2), (2, 4, 2), (2, 5, 2), (1, 3, 3), (2, 3, 3), (2, 4, 3), (2, 4, 4)\} \]
and corresponding set of gaps
\[ \{1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 19, 20, 23, 25, 26, 29, 38\} \]

We have described the semigroup \( W_{Q_n} \), and \( W_{P_n} \) as well, completely for \( n = 2 \) and \( n = 3 \) by giving all gaps. What happens if we increase \( n \) further? In this case we find only a partial description of the gaps. We consider an example.

**Example 3.5.19** Suppose \( q = 2 \). We study the semigroup \( W_{Q_4} \). Since the curve \( X_1 \) has genus 13, we need to find 13 gaps. Using Corollary 3.5.14 we find that the differential form
\[ x_1^{i_1} \cdots x_4^{i_4} \omega_{A_4} \]
is regular if and only if \((i_1, \ldots, i_4)\) is from the set
\[
\{(1, 1, 1, 1), (1, 2, 1, 1), (1, 3, 1, 1), (1, 3, 2, 1), (1, 4, 1, 1), (1, 4, 2, 1), (2, 1, 1, 1), (2, 1, 2, 1), (2, 1, 3, 1), (2, 1, 3, 2), (2, 1, 3, 3), (2, 2, 1, 1), (2, 2, 1, 2), (2, 2, 1, 3), (2, 2, 2, 1), (2, 2, 2, 2), (2, 2, 2, 3)\}

However the corresponding differential forms are not linearly independent. In fact a maximal linearly independent subset is obtained when making differential forms from the following set of 4-tuples
\[
\{(1, 1, 1, 1), (1, 1, 1, 2), (1, 1, 2, 1), (1, 2, 1, 1), (1, 2, 2, 1), (1, 3, 1, 1), (1, 3, 1, 2), (1, 3, 2, 1), (1, 3, 2, 2), (1, 3, 3, 1), (1, 3, 3, 2), (1, 3, 3, 3)\},
\]
with corresponding gaps
\[
\{1, 2, 3, 4, 5, 7, 9, 13, 17, 25\}.
\]
Hence, we have found 10 gaps.

The following table gives for \(n = 4\) and \(q \leq 17\) the number of gaps \(h\) one can find using the method we have described in this section. The total number of gaps \(g\) (which is the genus of the curve) is given in the third row.

<table>
<thead>
<tr>
<th>(q)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>11</th>
<th>13</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>(h)</td>
<td>10</td>
<td>50</td>
<td>176</td>
<td>421</td>
<td>1555</td>
<td>2611</td>
<td>4126</td>
<td>9010</td>
<td>17291</td>
<td>38965</td>
<td>49420</td>
</tr>
<tr>
<td>(g)</td>
<td>13</td>
<td>79</td>
<td>261</td>
<td>646</td>
<td>2493</td>
<td>4249</td>
<td>6796</td>
<td>15115</td>
<td>29394</td>
<td>67185</td>
<td>85528</td>
</tr>
</tbody>
</table>

We know that for \(n = 4\) the genus of the curve is given by
\[
g = q^4 + \frac{1}{2}q^3 - \frac{3}{2}q^2 + 1.
\]
However, for the calculated values of \(h\) we find
\[
h = \frac{1}{24}(13q^4 + 22q^3 - 25q^2 - 34q + 24),
\]
so it is reasonable to conjecture that this holds for all \(q\). If this is true we see that we find at least 13/24 part of the gaps. For \(n = 5\) we have the following table

<table>
<thead>
<tr>
<th>(q)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>11</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>(h)</td>
<td>22</td>
<td>160</td>
<td>595</td>
<td>1626</td>
<td>7415</td>
<td>13599</td>
<td>23288</td>
<td>58653</td>
<td>127462</td>
</tr>
<tr>
<td>(g)</td>
<td>33</td>
<td>280</td>
<td>1185</td>
<td>3576</td>
<td>18768</td>
<td>36225</td>
<td>64720</td>
<td>174120</td>
<td>397320</td>
</tr>
</tbody>
</table>
3.6 Gap-lifting in the Garcia-Stichtenoth towers

In this section we apply the technique of gap-lifting as explained in Proposition 3.2.8 to the first Garcia-Stichtenoth tower. In this way, we will obtain more information about the semigroup $W_P$. We will need to define the second Garcia-Stichtenoth tower $(S_n)_{n \geq 1}$.

**Definition 3.6.1** The second Garcia-Stichtenoth tower is defined inductively by $S_1 = F_q^*(\xi_1)$ and $S_{n+1} = S_n(\xi_{n+1})$, where

$$
\xi_{n+1}^q + \xi_{n+1} = \frac{\xi_n^q}{\xi_{n-1}^q + 1}.
$$

At first glance the first and second Garcia-Stichtenoth towers seem different, but they are related. The next proposition expresses this.

**Proposition 3.6.2** Denote by $F_n$ (respectively $S_n$) the $n$-th function field from the first (respectively second) Garcia-Stichtenoth tower. Then we have for $n \geq 1$

$$
S_n \subset F_{n+1}.
$$

Denote by $P'_n$ the unique place of $S_n$ having a pole at $\xi_1$. Then $P_{n+1}$ lies above $P'_n$ and $e_{P_{n+1}^*} = q + 1$.

**Proof:** We remarked before that the first Garcia-Stichtenoth tower can be defined as follows: $F_1 = F_q^*(x_1)$ and for $n \geq 1$ we have $F_{n+1} = F_n(z_{n+1})$ where

$$
z_{n+1}^q + z_{n+1} = x_{n+1}^q
$$

and

$$
z_{i+1} = x_{i+1} x_i.
$$

Using this, we see for $n \geq 2$

$$
z_{n+1}^q + z_{n+1} = \frac{z_{n+1}^q}{x_{n+1}} = \frac{z_n^q}{1 + z_n^{q-1}}.
$$

Hence the map $\phi : S_n \mapsto F_{n+1}$ defined by

$$
\phi(\xi_1, \ldots, \xi_n) = (z_2, \ldots, z_{n+1}),
$$

is well-defined and gives a monomorphism of function fields. In this way, we can write $F_{n+1} = S_n(x_1)$, with $x_1^{q-1} = z_2^q + z_2$. The point $P'_n \in P(S_n)$ is characterized by the fact that it has a pole at $\xi_1$ (see Lemma 3.3 in (Garcia and Stichtenoth 1996)). Hence any place in $P(F_{n+1})$ lying above it will have a pole at $x_1$ (use the fact that $\phi(\xi_1) = z_2$). However, because $P_{n+1}$ is the only place of $F_{n+1}$ having a pole at $x_1$, we conclude that the ramification is total. Hence $e_{P_n}^*P'_n = q + 1$. 

Remark 3.6.3 Note that the involution \( \iota_{n+1} \) of \( \mathcal{F}_{n+1} \) mentioned in Remark 3.5.10 induces an involution on \( \mathcal{S}_n \). In fact we have

\[
\iota_{n+1}(z_2, \ldots, z_{n+1}) = \left( \frac{1}{z_{n+1}}, \ldots, \frac{1}{z_2} \right).
\]

Hence the map \( \iota'_n : \mathcal{S}_n \rightarrow \mathcal{S}_n \) defined by

\[
\iota'_n(\xi_1, \ldots, \xi_n) = \left( \frac{1}{\xi_n}, \ldots, \frac{1}{\xi_1} \right)
\]
describes an involution of the function field \( \mathcal{S}_n \). This implies for example that the semigroup of the point \( Q_n \) (determined by the fact that it is a zero of the function \( \xi_n \)) is the same as the semigroup of the point \( P'_n \).

The reason we are interested in the second Garcia-Stichtenoth tower is the fact the the semigroup of the point \( P'_n \) has been calculated in (Pelikaan, Stichtenoth, and Torres 1998). We cite some of their results in the following theorem.

Theorem 3.6.4 Let \( P'_n \) be the unique place in the \( n \)-th function field \( \mathcal{S}_n \) from the second Garcia-Stichtenoth tower at which \( \xi_1 \) has a pole. The semigroup \( W_{P'_n} \) is equal to \( \Lambda_n \), where \( \Lambda_n \) is defined inductively by

\[
\Lambda_1 = \mathbb{N}_0
\]
and for \( n \leq 1 \)

\[
\Lambda_{n+1} = q \cdot \Lambda_n \cup \{ l \in \mathbb{Z} | l \geq c_{n+1} \}.
\]

where

\[
c_n = \begin{cases} 
q^n - q^{\frac{n+1}{2}} & \text{if } n \text{ is odd} \\
q^n - q^{\frac{n}{2}} & \text{if } n \text{ is even}
\end{cases}
\]

With this information we obtain the following proposition.

Proposition 3.6.5 Define

\[
G_n = \mathbb{N}_0 \setminus \Lambda_n,
\]
where \( \Lambda_n \) is as in the previous theorem. Then the elements in the set

\[
(q + 1)G_n
\]
are gaps at the point \( P_{n+1} \).

Proof: This follows directly from Corollary 3.2.9 and Proposition 3.6.2. \( \square \)
Example 3.6.6 Assume $q = 2$. In this case we have

$$G_3 = \{1, 2, 3\}$$

and hence we find for $P_4$ the set of gaps

$$\{3, 6, 9\}.$$ 

Of these we had not find 6 before, so we see that we have obtained one more gap, bringing the set of known gaps to

$$\{1, 2, 3, 4, 5, 6, 7, 9, 13, 17, 25\}.$$ 

In general we will find for $n = 4$ some more gaps, but it can be shown that $G_3$ has cardinality $(q^2 - 1)(q - 1)$. Hence for large $q$ this will probably not make a large difference, since we expect to find about $13q^4/24$ gaps with the methods described in the last section.
Chapter 4

Missing functions for plane curves of type II

Summary In this chapter we focus on the problem of finding the missing functions for plane curves of type II. In the case of non-degenerate plane curve of type II we find all these functions.

4.1 Introduction

Let \( \mathcal{X} \) be an algebraic curve over the finite field \( \mathbb{F}_q \). As mentioned before, the ring \( K_\infty(P) \) of functions regular outside the point \( P \) is important in the construction of codes. In practice, only a part of this ring is explicitly known. The rest of the ring is said to consist of “missing functions”.

Example 4.1.1 Consider the Klein quartic \( \mathcal{X} \) given by the equation

\[
X^3Y + Y^3 + X = 0,
\]

over the field \( \mathbb{F}_2 \). Denote by \( P \) the point \( [1 : 0 : 0] \) and by \( Q \) the point \( [0 : 1 : 0] \). In Example 3.1.3 we have seen that the functions

\[
x^iy^j
\]

are regular outside \( P \) if and only if \( 3i + j \geq 0 \) and \( -2i - 3j \geq 0 \). Further we have found for every number in the semigroup \( W_P \) a function that is regular outside \( P \) and has pole order at \( P \) equal to this number. Hence, in this case we have found a basis of the ring \( K_\infty(P) \) of functions regular outside \( P \).

Example 4.1.2 This example is taken from Chapter X, p. 914 of (Pless, Huffman, and Brualdi 1998). Let \( \mathcal{X} \) be the curve given by the equation
over a field $F_q$ of characteristic not equal to 2 or 3. This is a curve of type II with $a = 3$ and $b = 2$ and $c = 1$. Denote by $P$ the point $[0 : 1 : 0]$ of $\mathcal{X}$. The point $P$ is a singularity point, but only one point lies above it in the non-singular model. Since the characteristic is different from 2 and 3, we see that the curve is non-degenerate with respect to its Newton polygon. We conclude from Baker's theorem that the curve has genus 3. Further we deduce from Theorem 3.4.1 that the gaps of the semigroup $W_P$ are given by the set $\{1, 2, 4\}$. It follows that the semigroup $W_P$ is generated by the numbers 3 and 5 and 7. Hence, in order to describe the ring $K_\infty(P)$, it is sufficient to find three functions, regular outside $P$, such that their pole orders form the set $\{3, 5, 7\}$. It is not hard to see that the function $x^iy^j$ is regular outside $P$ if and only if $i \leq j$. Further this function has pole order $2i + 3j$ at $P$. Hence we have found a function having a pole of order 3 (respectively 5) at $P$, namely $y$ (respectively $xy$). But how do we find a function regular outside $P$ having pole order 7? In this example we can solve this by noting that the following functional identity on $\mathcal{X}$ holds.

$$(x + x^2y)^2 = -xy^4 - y^3 + xy + 1.$$ 

Since the right hand side of this equation is a linear combination of functions regular outside $P$, the left hand side is an element of $K_\infty(P)$ as well. We conclude that $x + x^2y \in K_\infty(P)$. Further note that the function $x + x^2y$ has pole order 7 at $P$.

In the previous example we found a missing function (using an ad hoc method). One would like to find the missing functions of a curve which gives rise to good codes. Especially the curves corresponding to the towers of function fields as given in (García and Stichtenoth 1995) and (García and Stichtenoth 1996) are suited for this purpose. For the first curves of these families the missing functions were found in (Voss and Hoholdt 1997). For the second tower of function fields, more missing functions were given in (Pelikan 1997), giving a complete description of the missing functions of the first three curves occurring in the second tower. We also mention the work in (Feng, Wu, and Rao 1998), where some missing functions are given. Instead of explicitly describing the missing functions, one could try to find an algorithm that gives these functions for a particular curve in a reasonable amount of time. This approach turned out to be a very fruitful one and gave a fast algorithm. See for example (Aleshnikov, Deodhar, Kumar, and Stichtenoth 1999), (Shum 2000) and (Shum, Aleshnikov, Kumar, and Stichtenoth 2000). Another algorithm was found independently in (Leonard 2000). This last algorithm has been used to corroborate some instances of the results mentioned in this chapter. The question of the missing functions of plane curves of type II was partially answered in Chapter X of (Piess, Huffman, and Brua 1998). There functions of the form $x^iy^j$ were considered. In the next section we will describe the ring $K_\infty(P)$ for non-degenerate plane curves of type II completely.
4.2 Plane curves of type II

Recall that a plane curve of type II is given by an equation of the form

\[ Y^{b+c} + uX^aY^c + G(X, Y) = 0. \]

Here \( a \) and \( b \) are relative prime numbers and \( u \) a non-zero constant from a perfect field \( F \). Further \( G \) satisfies \( \text{wdeg}(G) < ab \) and \( \text{deg}_X(G) < a \). The weighted degree \( \text{wdeg} \) is defined by \( \text{wdeg}(X) = b \) and \( \text{wdeg}(Y) = a \). The number \( c \) is assumed to be positive. The Newton diagram is depicted in Figure 3.4. Denote by \( P \) the point on the non-singular model corresponding to the edge \( \gamma \) connecting \( (0, b+c) \) and \( (a, c) \). Then \( v_P(X) = -b \) and \( v_P(Y) = -a \). In Theorem 3.4.1 all gaps of the semigroup \( W_P \) are described. Further in Proposition 3.4.4 generators are given of this semigroup. Our goal is to explicitly give a set of functions in \( K_\infty(P) \) whose pole orders at \( P \) will generate \( W_P \). First we give a lemma.

**Lemma 4.2.1** Let \( R \) be a domain and denote by \( Q(R) \) its field of fractions. Denote the normalization (or integral closure) of \( R \) by \( \tilde{R} \). Suppose that an elements \( q \in Q(R) \) and \( a_k \in \tilde{R} \), with \( 0 \leq k \leq n \), exist such that

\[ \sum_{k=0}^{n} a_k q^{n-k} = 0. \]

Then for all \( 0 \leq l < n \) the element

\[ q \sum_{k=0}^{l} a_k q^{l-k} \]

is an element of \( \tilde{R} \).

**Proof:** Write

\[ b_l = \sum_{k=0}^{l} a_k q^{l-k} \]

We will prove the theorem by induction on \( l \). First suppose \( l = 0 \). By multiplying the given relation by \( a_0^{n-1} \) we find the relation

\[ (a_0 q)^n + \sum_{k=1}^{n} a_k a_0^{k-1} (a_0 q)^{n-k} = 0. \]

Hence \( a_0 q \in \tilde{R} \).

Now suppose that the statement in the theorem is true for \( l - 1 \), where \( l \leq n \). We can rewrite the given relation as

\[ (b_{l-1} q + a_l) q^{n-l} + \sum_{k=l+1}^{n} a_k q^{n-k} = 0. \]
Note that

\[ b_t = b_{t-1}q + a_t \]

and hence by the induction hypothesis \( b_t \in \tilde{R} \). We found the relation

\[ b_t q^{n-l} + \sum_{k=l+1}^{n} a_k q^{n-i} = 0. \]

By a similar trick as before this implies \( b_t q \in \tilde{R} \). \( \square \)

The significance of this lemma lies in the fact that \( R \subset K_\infty(P) \) implies \( \tilde{R} \subset K_\infty(P) \). We will apply the above lemma to the case of a curve of type II. We will choose for \( R \) the ring \( \mathbb{F}[y] \) and for \( P \) the point on the non-singular model corresponding to \( [0 : 1 : 0] \). We will see later that the function \( y \) is regular outside \( P \) in all cases we will be working in. The following proposition gives some more information about \( K_\infty(P) \).

**Proposition 4.2.2** Let \( \mathcal{X} \) be a type II curve given by the equation

\[ Y^{b+c} + uX^aY^c + G(X, Y) = 0, \]

with the usual conditions on \( a, b, c, u \) and \( G \). Denote by \( P \) the point on the non-singular model of \( \mathcal{X} \) corresponding to \( [0 : 1 : 0] \) and write

\[ Y^{b+c} + uX^aY^c + G(X, Y) = \sum_{k=0}^{a} p_k(Y)X^{a-k}. \]

Assume that \( y \in K_\infty(P) \). Each of the functions

\[ x \sum_{k=0}^{l} p_k(y)x^{l-k} \]

\((0 \leq l < a)\) is regular outside \( P \) and has a pole at \( P \) of order \( ca + (1 + l)b \).

**Proof:** The fact that the functions mentioned above are regular outside \( P \) follows directly from Lemma 4.2.1 and the remarks following it. To compute the pole order at \( P \) note that for all monomial \( x^iy^j \) occurring in

\[ \sum_{k=0}^{l} p_k(y)x^{l-k} \]

we have

\[ v_P(x^iy^j) = -bi - aj \geq -bl - ca, \]

with equality only for the monomial \( x^iy^c \). \( \square \)
4.2 Plane curves of type II

Since \( y \in K_\infty(P) \) as well (by assumption), we can multiply the functions from the above Proposition by a power of \( y \) to find more elements of \( K_\infty(P) \). In this way we find functions in \( K_\infty(P) \) whose pole orders are of the form

\[(c + m)a + (1 + l)b,\]

where \( 0 \leq l < a \) and \( 0 \leq m \).

**Example 4.2.3** Let

\[ F(X, Y) = X^{13}Y^5 + Y^{17} + X^{12}Y^3 + X^9Y^2 + X^4(Y + Y^2) + 1. \]

The curve given by the equation \( F(X, Y) = 0 \) is a plane curve of type II with \( a = 13 \), \( b = 12 \) and \( c = 5 \). We know from the above that the following functions are elements of \( K_\infty(P) \).

\[
\begin{align*}
xy^5 & \\
x^2y^5 & + xy^3 \\
x^3y^5 & + x^2y^3 \\
x^4y^5 & + x^3y^3 \\
x^5y^5 & + x^4y^3 + xy^2 \\
x^6y^5 & + x^5y^3 + x^2y^2 \\
x^7y^5 & + x^6y^3 + x^3y^2 \\
x^8y^5 & + x^7y^3 + x^4y^2 \\
x^9y^5 & + x^8y^3 + x^5y^2 \\
x^{10}y^5 & + x^9y^3 + x^6y^2 + x(y + y^2) \\
x^{11}y^5 & + x^{10}y^3 + x^7y^2 + x^2(y + y^2) \\
x^{12}y^5 & + x^{11}y^3 + x^8y^2 + x^3(y + y^2) \\
x^{13}y^5 & + x^{12}y^3 + x^9y^2 + x^4(y + y^2)
\end{align*}
\]

Note that the last function is equal to \(-1 - y^{17}\). Using the (as yet unproved) fact that \( y \in K_\infty(P) \), we see that we can find a function regular outside \( P \) having pole order \( 12i + 13j \) at \( P \) as long as either \( i = 0 \) and \( j \geq 0 \) or \( i > 0 \) and \( j \geq 5 \). From the description of the semigroup given in Theorem 3.4.1, we see that we did not find a complete description of the set of regular functions. We are still missing functions with pole orders \( 12i + 13j \) with \((i, j)\) in the set

\[
\{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 3), (3, 4), (4, 3), (4, 4), (5, 4), (6, 4), (7, 4), (8, 4)\}.
\]

A reasonable guess would be that these functions can be obtained from the ones mentioned above by dividing by a suitable power of \( y \). Then the function corresponding to the pair \((4, 3)\) would be \((x^4y^5 + x^3y^4)/y^2 = x^4y^3 + x^3y^2\). In the same way the function corresponding to the pair \((7, 4)\) would be \(x^7y^4 + x^6y^3 + x^3y\). We
will prove in the following theorem that under some mild conditions on $F(X,Y)$, we can obtain in this way all missing functions from the functions mentioned in Proposition 4.2.2.

**Theorem 4.2.4** Let $\mathcal{X}$ be a curve of type $I\!I$ given by the equation

$$Y^{b+c} + uX^aY^c + G(X,Y) = 0,$$

with the usual conditions on $a$, $b$, $c$, $u$ and $G$. Denote by $P$ the point on the non-singular model of $\mathcal{X}$ corresponding to $[0 : 1 : 0]$ and write

$$Y^{b+c} + uX^aY^c + G(X,Y) = \sum_{k=0}^{a} p_k(Y)X^{a-k}.$$  

Suppose that the polynomial $F(X,Y)$ is non-degenerate with respect to its Newton polygon and that the curve $\mathcal{X}$ has no singularities outside the points $[0 : 0 : 1], [0 : 1 : 0], [1 : 0 : 0]$. Each of the functions

$$xy^l \sum_{k=0}^{l} p_k(y)x^{l-k}$$

$(0 \leq l < a)$ is regular outside $P$ and has a pole at $P$ of order $(c+i)a + (1+l)b$ as long as the point $(a-l-1,-i)$ lies beneath or on the boundary of the Newton polygon of $F(X,Y)$. Together with the function $y$, these functions form a basis of the ring $K_{\infty}(P)$.

**Proof:** We know that the line piece connecting $(0,b+c)$ and $(a,c)$ is an edge of the Newton polygon. Call this edge $\gamma$. Number the other edges from left to right. In this way, we find edges $\gamma_1, \ldots, \gamma_n$ for some $n$, having strictly increasing slopes. If we assume that the curve is non-degenerate with respect to its Newton polygon, we know that to every edge $\gamma_k$ correspond $l(\gamma_k)$ points on the non-singular model. Here $l$ stands for the arithmetic length. Suppose that $\gamma_k$ lies on the line given by the equation

$$a_k i + b_k j = c_k,$$

where $a_k$ and $b_k$ are relatively prime number and further suppose that the Newton polygon is contained in the half space

$$a_k i + b_k j \geq c_k.$$

Then we know from Proposition 3.3.1 that for any point $Q$ on the non-singular model corresponding to the edge $\gamma_k$ we have

$$v_Q(x) = a_k$$

and
4.2 Plane curves of type II

\[ v_Q(y) = b_k. \]

From the shape of the Newton polygon we see that all \( b_k \) are non-negative. This proves that \( y \in K_\infty(P) \). One further property of the numbers \( a_k \) and \( b_k \) is given by the equation

\[ a_k b_l - b_k a_l > 0, \]

if \( k < l \). This formula is reminiscent of the fact that the slopes of the edges \( \gamma_k \) increase as we go from left to right.

Fix some \( 0 \leq l < a \). We will now investigate the function

\[ S_l(x, y) = \sum_{k=0}^{l} p_k(y)x^{a-k}. \]

The function we claim is an element of \( K_\infty(P) \) can be written as

\[ x^{l-a+1}y^dS_l(x, y), \]

where \((a - l - 1, -j)\) is a point lying beneath the Newton polygon of \( F \). Define the edge \( \gamma_r \), depending on \( l \) in the following way. If the line \( i = a - l \) cuts one of the edges \( \gamma_1, \ldots, \gamma_n \) in two parts, define \( \gamma_r \) to be this edge. If the line goes through a vertex of the Newton polygon define \( \gamma_r \) to be the edge immediately left to this vertex. Further define \( \beta_{a-l} \) to be the point of intersection of \( \gamma_r \) and the line \( i = a - l \).

Suppose that \( Q \) is a point corresponding to the edge \( \gamma_p \). First assume that \( \rho \leq r \). In this case we see that

\[ v_Q(S_l(x, y)) \geq (a - l)a_p + \beta_{a-l}b_p \geq (a - l - 1)a_p + \beta_{a-l-1}b_p. \]

For the last inequality we use the fact that \( \gamma_p \) lies to the left of \( \gamma_r \). If \((a - l - 1, -j)\) is a point lying beneath the Newton polygon of \( F \), then we have \(-j \leq \beta_{a-l-1}\). Hence

\[ v_Q(x^{l-a+1}y^dS_l(x, y)) \geq (\beta_{a-l-1} + j)b_p \geq 0. \]

Now suppose that \( \rho \geq r \). Note that

\[ S_l(x, y) = \sum_{k=l+1}^{a} p_k(y)x^{a-k}, \]

Hence we have

\[ v - Q(S_l(x, y)) \geq (a - l - 1)a_p + \beta_{a-l-1}b_p \]

and from this we see as before that

\[ v_Q(x^{l-a+1}y^dS_l(x, y)) \geq 0. \]
The statement about pole orders at $P$ is clear. Comparing the pole orders we found with the description of the semigroup $W_P$ as given in Theorem 3.4.1 we see that we have found a basis for $K_\infty(P)$.
Chapter 5

Elliptic curves and large prime numbers

Summary In this chapter we investigate a connection between certain large prime numbers and elliptic curves. In particular we will describe two Mersenne-like families of prime numbers. This is joint work with J. M. Domnen.

5.1 Prime-generating elliptic curves

Let $E$ be an elliptic curve defined over a finite field $F_q$. We can also consider $E$ over the extension field $F_{q^e}$. More precisely, denote by $E(F_{q^e})$ the group of points of $E$ defined over the field $F_{q^e}$ and denote by $E_e$ the order of this group. The following theorem by Hasse and Weil gives some information about these numbers. Note that this theorem is a direct consequence of Theorem 1.2.2 mentioned in Chapter 1.

Theorem 5.1.1 (Hasse-Weil) There exists an algebraic integer $\alpha$, depending on the elliptic curve $E$, of length $\sqrt{q}$ such that

$$E_e = q^e + 1 - (\alpha^e + \overline{\alpha}^e).$$

For a proof see for example p.134 of (Silverman 1986). Note that $E(F_q)$ is a subgroup of $E(F_{q^e})$. Hence $E_1$ trivially divides $E_e$ for all $e$.

Definition 5.1.2 Let $E$ be an elliptic curve defined over a finite field $F_q$. We call $E$ a prime-generating elliptic curve if there exist infinitely many prime numbers of the form $E_e/E_1$.

It is, unfortunately, not known whether or not there exist prime-generating elliptic curves, but we will give some good candidates later. First we narrow down the possible values of $e$ for which $E_e/E_1$ can be prime.

91
**Proposition 5.1.3** Let $E$ be an elliptic curve defined over a finite field $\mathbb{F}_q$. Suppose that $E_c/E_1$ is a prime. Then $e$ is a prime or we are in one of the following two cases:

Case I: $q = 2$ and $e \in \{4, 6, 9\}$,
Case II: $q = 3$ and $e = 4$.

**Proof:** If $e$ has a non-trivial factor $f$, then $\mathbb{F}_{q^f}$ is a subfield of $\mathbb{F}_q$. Hence $E(\mathbb{F}_{q^f})$ is a subgroup of $E(\mathbb{F}_q)$ which implies that $E_f/E_1$ divides $E_c/E_1$. The only problem that is left is to show that, except for the two cases mentioned in the proposition, this is a non-trivial divisor.

Suppose that $e$ is not a prime, say that $e$ has a non-trivial divisor $f$. Since $f$ is a non-trivial divisor of $e$, we can assume that $\lfloor \sqrt{e} \rfloor \leq f \leq e/2$. Here $\lfloor \cdot \rfloor$ denotes the ceiling function. Hence if either $q \geq 3$ or $e \geq 5$, we have

$$E_f \leq (\sqrt{q^f} + 1)^2 \leq (\sqrt{q^{f/2}} + 1)^2 < (\sqrt{q} - 1)^2 \leq E_c.$$ 

For the first and the last inequality we used Hasse-Weil’s theorem. If $q = 2$ and $e = 4$ the strict inequality becomes an equality. So we see that for $q = 2$ either $E_2 < E_4$ or $E_2 = E_4 = 9$.

Now we will investigate if $E_f/E_1$ can be one. We know

$$E_1 \leq (\sqrt{q} + 1)^2 < (\sqrt{q} - 1)^2 \leq E_f,$$

whenever $q \geq 2$ and $e > 9$ or $q = 3$ and $e > 4$ or $q = 4$ and $e > 4$ or $q \geq 5$. Here we use $f \geq \lceil \sqrt{e} \rceil$. Note that for $q = 2$ and $e = 8$ strict inequality holds as well, since in this case $f = 4$.

So far we have proved that $E_f/E_1$ is a non-trivial divisor of $E_c/E_1$ whenever $q = 2$ and $e \not\in \{4, 6, 9\}$ or $q = 3$ and $e \not= 4$ or $q = 4$ and $e \not= 4$ or $q \geq 5$. We still need to exclude $q = 4$ and $e = 4$. What could go wrong is that for some curve $E_1 = E_2$ and $E_4/E_1$ is a prime number. It is not hard to see that in this case $\alpha = -2$ (from Hasse-Weil’s theorem) and hence $E_4/E_1 = 25$, which is not a prime. \hfill $\square$

**Remark 5.1.4** It is not hard do determine exactly for which curves $E$ the above exceptions occur. They are given below.

The first curve with this behaviour has Weierstrass equation

$$Y^2 + Y = X^3$$

and field of definition $\mathbb{F}_2$. In this case $E_4/E_1 = 3$.

Another curve defined over $\mathbb{F}_2$ giving rise to an exception has equation

$$Y^2 + Y = X^3 + X.$$ 

For this curve we have $E_1/E_4 = 5$, $E_6/E_1 = 13$ and $E_9/E_1 = 109$.

There is one more curve defined over $\mathbb{F}_2$ giving rise to an exception. It has Weierstrass equation
5.1 Prime-generating elliptic curves

\[ Y^2 + XY = X^3 + 1. \]

For this curve we have \( E_9/E_1 = 127. \)

The following curve defined over \( F_3 \) is the last curve giving rise to an exception. It has equation

\[ Y^2 = X^3 - X + 1. \]

For this curve \( E_4/E_1 = 13. \) The above list gives a complete description of the curves giving rise to the exceptions mentioned in Proposition 5.1.3.

**Remark 5.1.5** Note that it is not true that whenever \( e \) is a prime, the number \( E_e/E_1 \) is a prime as well. We are in a similar situation as in the case of the Mersenne numbers. These are numbers of the form \( 2^n - 1. \) It is not hard to show that such a number cannot be prime unless \( n \) is a prime. Is \( n \) is a prime, then the corresponding Mersenne number need not be prime. Another way to see that it is necessary for a Mersenne number to be prime, that the exponent \( n \) is prime is by considering the curve with equation

\[ Y^2 = X^3 + X^2 \]

over the field \( F_q. \) If we leave out the point \((0, 0)\) we find a group with an addition similar to that of an elliptic curve. The order \( E_e \) of this group over the extension field \( F_{q^e} \) is \( q^e - 1. \) Similar to Proposition 5.1.3 we can prove that the primality of \( E_e/E_1 \) implies that \( e \) is a prime. For \( q = 2 \) this gives the statement about Mersenne numbers, For a list of known Mersenne primes see http://www.utm.edu/research/primes/merSENNe.shtml \#known.

**Example 5.1.6** It is not true that any elliptic curve is prime-generating. If we look, for example, at the curve defined over \( F_4 \) having equation

\[ Y^2 + Y = X^3 + \theta, \]

where \( \theta \) is a generator for the group \( F_4. \) In this case \( E_1 = 1 \) and hence \( \alpha = 2. \) This implies

\[ \frac{E_e}{E_1} = (2^e - 1)^2. \]

**Example 5.1.7** On the other hand, if we look for example at the curve defined over \( F_2 \) given by the equation

\[ Y^2 + XY = X^3 + 1, \]

we find, using Mathematica and some hours of computer time, probable primes \( E_e/E_1 \) for \( e \) in the set
\{2, 5, 7, 9, 13, 19, 23, 41, 83, 97, 103, 107, 131, 233, 239, 277, 283, 349, 409, 571, 1249, 1913, 2221, 2647, 3169, 3527, 4349, 5333, 5903, 5923, 6701, 9127, 9829, 16187\}.

All \(e\) less than or equal to 17389 have been tried. Of course we used that for this curve \(e = 9\) or \(e\) is a prime. All prime candidates found pass the Mathematica function PrimeQ. If this function gives as outcome “False”, it is certain that the tested number is not a prime; if however its outcome is “True”, then it is likely, but not certain that the number is a prime. This explains why we said that we found probable primes. Note that for this curve \(E_1 = 4\) and hence \(\alpha = -1/2 + i\sqrt{7}/2\). This certainly looks like a promising candidate for a prime-generating elliptic curve. We conjecture that any elliptic curve for which an \(e\) exists such that \(E_e/E_1\) is prime, is in fact prime-generating.

**Example 5.1.8** As another example of a potential prime-generating elliptic curve we consider the curve defined over \(\mathbb{F}_2\) with Weierstrass equation

\[Y^2 + Y = X^3 + X.
\]

For this curve we have \(E_1 = 5\). Further \(\alpha = -1 + i\). It is not hard to see that for any odd prime \(e\) we have

\[E_e = 2^e + \left(\frac{2}{e}\right) 2^{(e+1)/2} + 1.
\]

Using Mathematica we find that for \(e\) in the following set the number \(E_e/E_1\) is probably a prime:

\[\{4, 5, 6, 7, 9, 11, 13, 17, 29, 43, 53, 89, 283, 557, 563, 613, 691, 1223, 2731, 5147, 5323, 5479, 9533, 10771, 11257, 11519, 12583\}.

Again we tried all \(e\) less than 17389.

**Example 5.1.9** Consider the elliptic curve defined over \(\mathbb{F}_2\) given by the Weierstrass equation

\[Y^2 + XY = X^3 + X + 1.
\]

For this curve we have \(E_1 = 2\) and hence \(\alpha = 1/2 + i\sqrt{7}/2\). For the following \(e\) we find probable primes of the form \(E_e/E_1\):

\[\{2, 3, 5, 7, 11, 17, 19, 23, 101, 107, 109, 113, 163, 283, 311, 331, 347, 359, 701, 1153, 1597, 1621, 2063, 2437, 2909, 3319, 6011, 12829\}.

This list is complete up to 17389.
Example 5.1.10 We consider in this example the elliptic curve with Weierstrass equation

\[ Y^2 + Y = X^3. \]

Then \( E_1 = 3 \) and \( \alpha = i\sqrt{2} \). The number \( E_\alpha/E_1 \) is probably a prime for the \( e \) in the following set

\[ \{2, 3, 4, 5, 7, 11, 13, 17, 19, 23, 31, 43, 61, 79, 101, 127, 167, 191, 199, 313, 347, 701, 1709, 2617, 3539, 5807, 10501, 10691, 11279, 12391, 14479 \} \]

As before this list is complete up to 17389.

In the next section we will see two more examples of Mersenne like families of primes. One of the nice things about Mersenne prime numbers is that there exists a very efficient primality test for them. See for example p.92 of (Ribenboim 1996) or p.27 of (Bressoud 1989). In the next section we will find a similar primality test for the numbers \( E_\alpha/E_1 \) for certain elliptic curves.

5.2 A primality test for certain elliptic curves

Loosely speaking it is fair to say that it is possible to find an efficient way of testing whether or not a number \( N \) is prime if either \( N + 1 \) or \( N - 1 \) can be factored. In practice it is not necessary to be able to factor \( N \pm 1 \) completely. The numbers we want to test have the form \( E_\alpha/E_1 \), where \( E_1 \) is some small number and \( E_\alpha = q^e + 1 - (\alpha^e + \beta^e) \) for some algebraic number \( \alpha \) of norm \( \sqrt{q} \). We see that in general there is no hope to factor \( E_\alpha/E_1 \pm 1 \), since all structure of the number \( E_\alpha \) is lost by the division by \( E_1 \). However for some elliptic curves \( E_1 = 1 \). We will classify all elliptic curves with \( E_1 = 1 \) in the following proposition.

Proposition 5.2.1 Let \( \mathcal{E} \) be an elliptic curve defined over \( \mathbb{F}_q \) with the property that \( E_1 = 1 \). Then we are in one of the following three cases.

Case I: \( q = 2 \) and the curve \( \mathcal{E} \) is isomorphic to the curve with Weierstrass equation

\[ Y^2 + Y = X^3 + X + 1, \]

Case II: \( q = 3 \) and the curve \( \mathcal{E} \) is isomorphic to the curve with Weierstrass equation

\[ Y^2 = X^3 - X - 1, \]

Case III: \( q = 4 \) and the curve \( \mathcal{E} \) is isomorphic to the curve with Weierstrass equation

\[ Y^2 = X^3 + \zeta, \]

with \( \zeta \) a generator of the multiplicative group \( \mathbb{F}_4^* \).
Proof: We will first determine the possible Frobenius eigenvalues $\alpha$. Write $\alpha = a + bi$, with $a$ and $b$ real numbers. Then we have $a^2 + b^2 = q$ by Hasse-Weil’s theorem. Further the assumption $E_1 = 1$ implies the equation $a = q/2$. Hence

$$\alpha = q/2 \pm i \sqrt{q - q^2/4}$$

and apparently we should demand $q - q^2/4 \geq 0$. This implies $q \in \{2, 3, 4\}$. This gives rise to the three cases mentioned in the proposition. \(\square\)

Note that for any of the three cases mentioned above we have $\alpha + \overline{\alpha} = q$, so the three curves are supersingular. We now determine the numbers $E_c$ for these three cases.

**Proposition 5.2.2** Let $E$ be the elliptic curve defined over $\mathbb{F}_2$ with Weierstrass equation

$$Y^2 + Y = X^3 + X + 1.$$

Then a Frobenius eigenvalue is given by $1 + i$. Further we have

$$E_c = \begin{cases} 2^c - 2^{c/2} + 1 & \text{if } e \equiv 0 \pmod{8}, \\ 2^c - 2^{(c+1)/2} + 1 & \text{if } e \equiv 1, 7 \pmod{8}, \\ 2^c + 1 & \text{if } e \equiv 2, 6 \pmod{8}, \\ 2^c + 2^{(c+1)/2} + 1 & \text{if } e \equiv 3, 5 \pmod{8}, \\ 2^c + 2^{c/2} + 1 & \text{if } e \equiv 4 \pmod{8}. \end{cases}$$

Let $E$ be the elliptic curve defined over $\mathbb{F}_2$ with Weierstrass equation

$$Y^2 = X^3 - X - 1.$$

Then a Frobenius eigenvalue is given by $3/2 + i\sqrt{3}/2$. Further we have

$$E_c = \begin{cases} 3^c - 3^{c/2} + 1 & \text{if } e \equiv 0 \pmod{12}, \\ 3^c - 3^{(c+1)/2} + 1 & \text{if } e \equiv 1, 11 \pmod{12}, \\ 3^c - 3^{c/2} + 1 & \text{if } e \equiv 2, 10 \pmod{12}, \\ 3^c + 1 & \text{if } e \equiv 3 \pmod{6}, \\ 3^c + 3^{c/2} + 1 & \text{if } e \equiv 4, 8 \pmod{12}, \\ 3^c + 3^{(c+1)/2} + 1 & \text{if } e \equiv 5, 7 \pmod{12}, \\ 3^c + 3^{c/2} + 1 & \text{if } e \equiv 6 \pmod{12}. \end{cases}$$

Let $E$ be the elliptic curve defined over $\mathbb{F}_2$ with Weierstrass equation

$$Y^2 = X^3 + \zeta,$$

where $\zeta$ generates the group $\mathbb{F}^*_2$. Then a Frobenius eigenvalue is given by $2$. Further we have

$$E_c = (2^c - 1)^2.$$
5.2 A primality test for certain elliptic curves

Proof: The results follow from a straightforward calculation. □

We see that the third curve cannot give new primes, since \( E_e \) is the square of a Mersemne number for any \( e \). For the first two curves we state the following corollary.

**Corollary 5.2.3** Let \( e \) be a prime not equal to 2 or 3. We have

\[
E_e = 2^e - \left( \frac{2}{e} \right) 2^{(e+1)/2} + 1
\]

for the first curve in Proposition 5.2.2. For the second curve in Proposition 5.2.2 we find

\[
E_e = 3^e - \left( \frac{3}{e} \right) 3^{(e+1)/2} + 1.
\]

Here \( \left( \frac{2}{e} \right) \) and \( \left( \frac{3}{e} \right) \) denote Legendre symbols.

We see that for these two curves we can partially factor \( E_e - 1 \). We now quote the theory necessary to describe the primality test for these numbers. For a proof see for example p.51 of (Ribenboim 1996).

**Theorem 5.2.4 (Pocklington)** Let \( n \) be a natural number and suppose that \( n - 1 = FR \), where \( F \) has known factorization

\[
F = p_1^{r_1} \cdots p_r^{r_r},
\]

for some primes \( p_1, \ldots, p_r \) and where \( R \) is relatively prime to \( F \). Suppose that for each \( i \) from 1 to \( r \) there exists a number \( a_i \) such that

\[
a_i^{n-1} \equiv 1 \pmod{n}
\]

and

\[
\gcd \left( a_i^{n-1}, n \right) = 1.
\]

Then for any divisor \( d \) of \( n \) we have

\[
d \equiv 1 \pmod{F}.
\]

Hence, if in addition \( F > \sqrt{n} \), then \( n \) is a prime.

**Corollary 5.2.5 (Proth)** Let \( n \) be an odd natural number and suppose that

\[
n - 1 = 2^e R,
\]

where \( R \) is odd and less than \( \sqrt{n} \). If a number \( a \) exists such that

\[
a^{\frac{n-1}{2^e}} \equiv -1 \pmod{n},
\]

then \( n \) is prime.
This leads immediately to a primality test for the numbers \(2^e - \left(\frac{2}{e}\right) 2^{(e+1)/2} + 1\) as we show in the following theorem.

**Theorem 5.2.6** Let \(e\) be an odd prime and define \(E_e = 2^e - \left(\frac{2}{e}\right) 2^{(e+1)/2} + 1\). Define

\[
a = \begin{cases} 
3 & \text{if } e \equiv 5, 7 \pmod{8}, \\
5 & \text{if } e \equiv 3 \pmod{8}, \\
2^n + 1 & \text{if } e \equiv 2^n + 1 \pmod{2^{n+2}}.
\end{cases}
\]

Here \(n \geq 2\). The number \(E_e\) is a prime if and only if

\[a^{(E_e-1)/2} \equiv -1 \pmod{E_e}.
\]

**Proof:** The implication from right to left follows directly from Proth’s primality test. For the other implication use the properties of Jacobi symbols. \(\square\)

A C-implemention of this test has revealed that the number \(2^e - \left(\frac{2}{e}\right) 2^{(e+1)/2} + 1\) is a prime for \(e\) in the set 
\[
\{2, 3, 5, 7, 11, 19, 29, 47, 73, 79, 113, 151, 157, 163, 167, 239, 241, 283, 353, 367, 379, 457, 997, 1367, 3041, 10141, 14699, 27529, 49207, 77291, 85237 \}.
\]

While we were looking for larger \(e\) we found out that someone else, namely M. Oakes, was looking for primes of this form as well. He found, using practically the same primality test, that \(E_e\) is prime for \(e\in\{106693, 160423, 203789\}\) as well. We have checked that for \(e \leq 188369\) this list is complete. As far as we know no published material in the literature exists about these primes.

Now we turn our attention to primes of the form \(3^e - \left(\frac{3}{e}\right) 3^{(e+1)/2} + 1\).

**Theorem 5.2.7** Define \(E_e = 3^e - \left(\frac{3}{e}\right) 3^{(e+1)/2} + 1\). The number \(E_e\) is a prime if and only if

\[2^{(E_e-1)/3} \equiv -3^e \pmod{E_e}
\]

or

\[2^{(E_e-1)/3} \equiv 3^e - 1 \pmod{E_e}.
\]

**Proof:** The if part follows directly from Pocklington’s theorem. Now we look at the only if part. Suppose that \(E_e\) is a prime. Note that if \(E_e\) is prime the solutions of the equation \(x^3 \equiv 1 \pmod{E_e}\) are given by \(1\) and \(-3^e\) and \(3^e + 1\). Hence we need to show that \(2\) is not a cubic residue modulo \(E_e\). Define \(\omega = -1/2 + i\sqrt{3}/2\) and \(\alpha = 3/2 + i\sqrt{3}/2\). Note that \(\alpha\) is a Frobenius eigenvalue of the elliptic curve over \(F_3\) having Weierstrass equation \(Y^2 = X^3 - X - 1\), which is the curve the numbers \(E_e\) come from. Note that \(\alpha = 2 + \omega\). Using the cubic reciprocity law in the ring \(\mathbb{Z}[\omega]\) (see, for example, pages 78-80 of (Cox 1989)) we find that \(2\) is never a cubic residue modulo \(E_e\), still assuming that \(E_e\) is a prime. The main fact we use, is that over \(\mathbb{Z}[\omega]\) we have the factorization \(E_e = (\alpha^p - 1)(\overline{\alpha^p} - 1)\). \(\square\)
Using a C-program we found that $E_e$ is a prime for $e$ in the set
\{5, 7, 11, 17, 19, 79, 163, 193, 239, 317, 353, 659, 709, 1049, 1103, 1759, 2029, 5153, 7541, 9049, 10453, 23743\}.

This list is complete through 115000. We did not find any reference in the literature for these prime numbers. Nor did, as far as we know, anyone look at these primes before.

5.3 The Wagstaff conjecture

In this section we generalize the Wagstaff conjecture for the Mersenne numbers to more general families of numbers arising from elliptic curves. Recently we learned that this was done independently in (Koblitz 2001) as well. For the Mersenne primes the following conjecture exists (see (Wagstaff 1983)).

**Conjecture 5.3.1 (Wagstaff)** The number of Mersenne primes less than or equal to $x$ is approximately $(e^\gamma / \log 2) \log \log x$. (Here $\gamma$ is Euler’s constant).

The expected number of Mersenne primes $2^p - 1$ with $p$ between $n$ and $2n$ is approximately $e^\gamma$.

The probability that $2^p - 1$ is prime is approximately $(e^\gamma \log ap)/p \log 2$ where $a = 2$ if $p \equiv 3 \pmod{4}$ and $a = 6$ if $p \equiv 1 \pmod{4}$.

There exists an heuristic argument in favour for this conjecture. Since we will extend this conjecture to more general numbers, we will describe the heuristic here. It has been taken from the internet. For the following and for more information see http://www.utm.edu/research/primes/mersenne/heuristic.html. First we observe that the probability that a number $2^k - 1$ is a prime can be estimated as follows. In the first place note that $e$ should be a prime. By the prime number theorem we see that the probability that a random number $M$ is a prime is approximately $1/\log M$, with $\log$ the natural logarithm. It is not hard to prove that any divisor $d$ of $2^p - 1$ satisfies $d \equiv 1 \pmod{ak}$, where $a = 2$ if $p \equiv 3 \pmod{4}$ and $a = 6$ if $p \equiv 1 \pmod{4}$. Hence we find the following expression for the probability that $2^k - 1$ is a prime:

$$\frac{1}{\log 2^k - 1} \prod_{\substack{p < ak + 1, \\ p \text{ a prime}}} \frac{p}{p - 1}.$$  

To estimate this probability we use Merten’s theorem (see for example (Bach and Shallit 1996)) which states that

$$\lim_{n \to \infty} \frac{1}{\log n} \prod_{i=1}^{n} \frac{p_i}{p_i - 1} = e^\gamma.$$
Here \( p_i \) denotes the \( i \)-th prime. We find the following estimate for the probability that \( 2^k - 1 \) is a prime:

\[
\frac{e^\gamma \log ak}{k \log 2}
\]

However, we have not taken into account yet that \( k \) should be a prime. This gives the following estimate for the probability that \( 2^k - 1 \) is a prime:

\[
\frac{e^\gamma}{k \log 2},
\]

where we neglected the \( a_i \) since we assume that \( k \) is big. Using this estimate, the other parts from Wagstaff’s conjecture follow immediately. For example the number of Mersenne primes less than a given constant \( x \) is approximately

\[
\sum_{k=1}^{(\log x/\log 2)} \frac{e^\gamma}{k \log 2} \approx \frac{e^\gamma}{\log 2} \log (\log x)/\log 2 \approx \frac{e^\gamma}{\log 2} \log (\log x),
\]

We see that, roughly speaking, the Wagstaff conjecture is based on the following observations. In the first place \( 2^n - 1 \) can be a prime only if \( n \) is a prime and in the second place if \( n \) is a prime an \( d \) divides \( 2^n - 1 \) then \( d \equiv 1 \pmod{n} \). Now let \( E \) be any elliptic curve. We have already seen that \( E_{e}/E_{1} \) can be a prime only if \( e \) is a prime (with the exception of some small \( e \) for some special curves) in Proposition 5.1.3. We now prove the analog of the second observation.

**Theorem 5.3.2** Let \( E \) be an elliptic curve defined over the finite field \( \mathbb{F}_q \). Suppose that \( e \) is a prime and that \( l \) is a prime divisor of the number \( E_{e}/E_{1} \). Then we have

\[
l \equiv 1 \pmod{e},
\]

or

\[
l \equiv -1 \pmod{e} \text{ and } l|(q - 1),
\]

or

\[
l = e \text{ and } l|E_1.
\]

**Proof:** Denote by \( E[l] \) the \( l \)-torsion of the curve \( E \) over the algebraic closure of \( \mathbb{F}_q \). Since \( l \) divides \( E_{e}/E_{1} \) we see that the intersection \( E[l] \cap E(\mathbb{F}_q) \) is not trivial. The Galois group \( \text{Gal}(\mathbb{F}_q'/\mathbb{F}_q) \) acts on this intersection. Since \( e \) is a prime number, we see that the orbits under this action either consist of one or \( e \) points. If an orbit consists of a single point \( P \), this means that \( P \) is defined over \( \mathbb{F}_q \). Hence the number of orbits of size one equals the cardinality of the group \( E[l] \cap E(\mathbb{F}_q) \). If \( l \) does not divide \( E_{1} \) this intersection is trivial. We have in this case

\[
e \mid \#(E[l] \cap E(\mathbb{F}_q)) - 1.
\]

There are two cases. We can have \( E[l] \cap E(\mathbb{F}_q) \cong \mathbb{Z}/l\mathbb{Z} \). Then we see that \( e \) divides \( l - 1 \) and hence \( l \equiv 1 \pmod{e} \). We also can have \( E[l] \cap E(\mathbb{F}_q) \cong \mathbb{Z}/l\mathbb{Z} \times \mathbb{Z}/l\mathbb{Z} \). Hence
we find $l \equiv \pm 1 \pmod{e}$. However, in this case using the Weil-pairing one can show that $\mathbb{Z}/l\mathbb{Z}$ is (isomorphic to) a subgroup of $\mathbb{F}_{q^e}^*$ (see for example p. 98 (Silverman 1986)). Hence we see $q^e \equiv 1 \pmod{l}$ from which we deduce $l \equiv 1 \pmod{e}$ or $l \equiv -1 \pmod{e}$.

If $l|E_1$ we can reason in a similar way. Denote by $\mathcal{E}[l^{\infty}]$ the group of points defined over the algebraic closure of order a power of $l$. Instead of using the group $\mathcal{E}[l] \cap \mathcal{E}(\mathbb{F}_{q^e})$ as above, we now look at the group $\{P \in \mathcal{E}(\mathbb{F}_{q^e}) \mid lP \in \mathcal{E}(\mathbb{F}_q) \cap \mathcal{E}[l^{\infty}]\}$. The elements of this group fixed by the Galois action form the group $\mathcal{E}(\mathbb{F}_q) \cap \mathcal{E}[l^{\infty}]$. Reasoning as before we see that

$$e|\#\{P \in \mathcal{E}(\mathbb{F}_{q^e}) \mid lP \in \mathcal{E}(\mathbb{F}_q) \cap \mathcal{E}[l^{\infty}]\} - \#\mathcal{E}(\mathbb{F}_q) \cap \mathcal{E}[l^{\infty}] .$$

As before we get that $l \mid q - 1$ and $l \equiv -1 \pmod{e}$ or $l \equiv 1 \pmod{e}$ as well as one more possibility, namely $l = e$. This concludes the proof. □

When trying to determine whether or not a number $E_c/E_1$ is a prime, one usually tries to find a factor prior to doing a primality test. The above theorem helps a lot when trying small divisors or applying the $(p - 1)$-factoring algorithm. For the two special families of numbers we found a primality test for before, we find the following proposition which is reminiscent of the above theorem, but a bit stronger.

**Proposition 5.3.3** Let $e$ be an odd prime larger than 3. Suppose that $l$ is a prime divisor of the number $2^e - (\frac{2}{3})^2 + 1$ (respectively $3^e - (\frac{2}{3})^3 + 1$). Then $l \equiv 1 \pmod{4e}$ or $l = 5$ (respectively $l \equiv 1 \pmod{6p}$ or $l = 7$).

**Proof:** Note that $2^e - (\frac{2}{3})^2 + 1$ is a divisor of $2^{2e} + 1$. This implies that $2^{2e} \equiv -1 \pmod{l}$. Hence the multiplicative order of 2 in the group $\mathbb{F}_l^*$ is a divisor of $4e$. This implies by some manipulation of the formulas that $l = 5$ or the multiplicative order of 2 equals $4e$ which implies $l \equiv 1 \pmod{4e}$. For the other half of the proposition note that $3^e - (\frac{2}{3})^3 + 1$ is a divisor of $3^{2e} + 1$ and proceed similarly. □

We see that we can extend the Wagstaff conjecture to the sequence of numbers $E_c/E_1$ for a given elliptic curve $\mathcal{E}$. More precisely we state the following analog of Wagstaff's conjecture.

**Conjecture 5.3.4** Let $\mathcal{E}$ be a prime-generating elliptic curve defined over the field $\mathbb{F}_q$. Denote by $\gamma$ Euler’s constant and the exponential function with groundnumber $e$ by exp. We have the following

The number of primes of the form $E_k/E_1$ less than or equal to $x$ is approximately

$$(\exp(\gamma)/\log q) \log \log x .$$

The expected number of primes of the form $E_k/E_1$ with $k$ between $n$ and $qn$ is approximately $\exp(\gamma)$. The probability that $E_k/E_1$ is prime is approximately
\[ \exp(\gamma) \log \frac{ak}{k \log q}, \]

where \( a \) depends on the specific choice of \( \mathcal{E} \), but generically \( a = 2 \).

For an elliptic curve defined over \( \mathbb{F}_2 \) there are exactly five \( \mathbb{F}_2 \) isomorphism classes. They correspond to the number of points the elliptic curve has defined over \( \mathbb{F}_2 \). For, when defined over \( \mathbb{F}_2 \), it can have 1 or 2 or 3 or 4 or 5 \( \mathbb{F}_2 \)-rational points. The next five examples give an idea of how good Wagstaff’s conjecture is for these curves. The conjecture seems to hold, but of course we have only a very limited amount of data.

**Example 5.3.5** Consider the elliptic curve discussed in Example 5.1.7. Define \( e_n \) to be the \( n \)-th number such that \( E_{e_n}/E_1 \) is a prime. According to Wagstaff’s conjecture the plot of \( n \) against \( \log_2(\log_2(E_{e_n}/E_1)) \) should lie on a straight line with slope \( \exp(-\gamma) \). In the following picture we have drawn this line as well as plotted the actual data. The values on the line are denoted by diamonds, while the actual data are denoted by stars.

![Figure 5.1: Example 5.3.5](image)

**Example 5.3.6** For the elliptic curve mentioned in Example 5.1.8 we find the following picture
Example 5.3.7 In this example we look at the curve having Weierstrass equation

\[ Y^2 + Y = X^3 + X + 1. \]

Recall that for this curve we have an efficient primality test. The picture looks as follows

Example 5.3.8 In this example we return to the elliptic curve mentioned in Example 5.1.9. The corresponding picture is
Example 5.3.9 Finally we look again at the curve mentioned in Example 5.1.10. For this curve we found the following picture

5.4 Jacobians of prime order

In the previous section we found two interesting sequences of numbers for which an efficient primality test exists. They arose naturally by considering certain elliptic curves having only only point defined over the field of definition ($E_1 = 1$). We now
5.4 Jacobians of prime order

look at Jacobians of curves. First we need a way to be able to count the number of points on a Jacobian. It turns out that Theorem 5.1.1 can be generalized.

**Theorem 5.4.1** Let $C$ be a curve of genus $g$ defined over the field $\mathbb{F}_q$ and let $J_C$ be its Jacobian. Further define $J_0 = \# J_C(\mathbb{F}_q)$. There exist, depending on $C$, complex numbers $\alpha_1, \ldots, \alpha_g$ of length $\sqrt{q}$ such that

$$J_0 = \prod_{i=1}^{g}(1 - \alpha_i \overline{\alpha_i}).$$

It is customary instead of giving the $\alpha_i$ explicitly, to give a polynomial of degree $2g$ which has as zeros exactly the $\alpha_i$ and $\overline{\alpha_i}$. The reciprocal of this polynomial is called the $L$-polynomial. Inspired by the elliptic curve case we restrict ourselves to Jacobians which have only one point defined over the field of definition. All curves giving rise to Jacobians having only one point defined over the ground field, have been classified. The following theorem is from (Leitzel, Madan, and Queen 1975). It is a generalization of Proposition 5.2.1. The theorem is posed in the language of function fields. The cardinality of the Jacobian over the ground field is called the class number.

**Theorem 5.4.2** Up to isomorphism, the following seven fields are the only congruence fields $\mathbb{F}$ (with $\mathbb{F}$ as field of definition) which have class number one and genus $g$ different from zero.

$F = \mathbb{F}_2$,
$g = 1$,
*defining equation*: $Y^2 + Y = X^3 + X + 1$
and $L(T) = 1 - 2T + 2T^2$,

$F = \mathbb{F}_2$,
$g = 2$, *defining equation*: $Y^2 + Y = X^5 + X^3 + 1$
and $L(T) = 1 - 2T + 2T^2 - 4T^4 + 4T^4$,

$F = \mathbb{F}_2$,
$g = 2$,
*defining equation*: $Y^2 + Y = (X^3 + X^2 + 1)/(X^3 + X + 1)$
and $L(T) = 1 - 3T + 5T^2 - 6T^3 + 4T^4$,

$F = \mathbb{F}_2$,
$g = 3$,
*defining equation*: $Y^4 + XY^3 + (X^2 + X)Y^2 + (X^3 + 1)Y + (X^4 + X + 1) = 0$
and $L(T) = 1 - 3T + 2T^2 + T^3 + 4T^4 - 12T^5 + 8T^6$,

$F = \mathbb{F}_2$,
$g = 3$,
defining equation: \( Y^4 + (X^3 + X + 1)Y + (X^4 + X + 1) = 0 \)
and \( L(T) = 1 - 3T + 3T^2 - 2T^3 + 6T^4 - 12T^5 + 8T^6 \),

\[ F = F_3, \]
\[ g = 1, \]
defining equation: \( Y^2 = X^3 - X - 1 \)
and \( L(T) = 1 - 3T + 3T^2 \),

\[ F = F_4, \]
\[ g = 1, \]
defining equation: \( Y^2 = X^3 + \theta \)
and \( L(T) = 1 - 4T + 4T^2 \),
where \( \theta \) generates the multiplicative group of \( F_4 \).

Note that case one, six and seven are the ones mentioned in Proposition 5.2.1. Of the other cases only case two gives rise to numbers \( n \) such that \( n - 1 \) can easily be partially factored. We will establish some facts about this Jacobian in the following proposition.

**Proposition 5.4.3** Let \( C \) be the curve defined over \( F_2 \) given by the equation

\[ Y^2 + Y = X^5 + X^3 + 1. \]

Then (notation being as in Theorem 5.4.1) we have

\[ \alpha_1 = \frac{1 + \sqrt{3}}{2} + i \frac{1 - \sqrt{3}}{2} \]

and

\[ \alpha_2 = \frac{1 - \sqrt{3}}{2} + i \frac{1 + \sqrt{3}}{2}. \]

Further, supposing that \( e \) is a prime not equal to 2 or 3, we have

\[ \prod_{e} = 2^{(e+1)/2} \left( 2^{(e-1)/2} - \left( \frac{2}{e} \right) \right) + 1. \]

**Proof:** To show that the \( \alpha_i \) are reciprocal to the zeros of the \( L \)-polynomial is a matter of straightforward checking. For the last statement note that \( (\alpha_i / \sqrt{3})^{12} = -1 \). This means that the powers of \( \alpha_i \) can be calculated explicitly from which we can obtain an explicit formula for \( \prod_{e} \). \( \Box \)

**Corollary 5.4.4** Let the notation be as in the above theorem. Any prime divisor \( l \) of \( \prod_{e} \) satisfies \( l \equiv 1 \pmod{12e} \) or \( l = 5 \) or \( l = 13 \).
5.4 Jacobians of prime order

**Proof:** This follows from the fact that \( J_e \) divides \( 2^{6e} + 1 \). Explicitly

\[
2^{6e+1} = J_e (2^{2e} + 1) \left( 2^{(e+1)/2} (2^e + 1) \left( 2^{(e-1)/2} + \left( \frac{2}{e} \right) \right) + 1 \right).
\]

We see that we can factor \( J_e - 1 = 2^{(e+1)/2} 3 \cdot R \), where \( R \) denotes the non-factored remainder. Unfortunately, this not enough to be able to apply Pocklington’s theorem to obtain a simple primitivity test. For that we need to factor half of \( J_e - 1 \). There exists a refinement of Pocklington’s theorem for which we only have to factor one third (see for example (Maurer 1995)). We do get the following result.

**Proposition 5.4.5** Let \( e \) be an odd prime not equal to 3 and let

\[
J_e = 2^{(e+1)/2} (2^e + 1) (2^{(e-1)/2} - \left( \frac{2}{e} \right)) + 1.
\]

Suppose that there exists a number \( a \) such that

\[
a^{(J_e - 1)/2} \equiv -1 \pmod{J_e},
\]

then \( J_e \) is the product of at most two primes.

**Proof:** Directly from Pocklington’s theorem, we see that any prime divisor of \( J_e \) is of the form \( 2^{(e+1)/2} k + 1 \) for some natural number \( k \). By looking at the size of these divisors, we see that \( J_e \) can be at most the product of three primes. Assume that this is the case, say

\[
J_p = (2^{(e+1)/2} k_1 + 1) (2^{(e+1)/2} k_2 + 1) (2^{(e+1)/2} k_3 + 1),
\]

for some positive \( k_i \). Then we find in the first place

\[
k_1 + k_2 + k_3 \equiv - \left( \frac{2}{e} \right) \pmod{2^{(e-1)/2}},
\]

but in the second place \( k_1 k_2 k_3 \leq 2^{(e-3)/2} \) and hence

\[
k_1 + k_2 + k_3 \leq 2 + 2^{(e-3)/2} < 2^{(e-1)/2}.
\]

From this we see that

\[
k_1 + k_2 + k_3 = - \left( \frac{2}{e} \right).
\]

This gives a contradiction. \( \square \)

We see that for the numbers \( J_e \) in the above proposition we do not have a primality test, but we do have a test showing that the number is the product of at most two
primes. A computer search revealed that for the following values of $e$ the number $J_e$ is the product of at most two primes:

$$\{5, 7, 11, 19, 23, 31, 239, 757, 859, 1723, 6079\}$$

For some small values of this list it is still possible to factor $J_e - 1$ further using a computer. This enables one to use Pocklington’s theorem to prove primality of $J_e$. Up till $e = 31$ this is trivial. We state some results for larger $e$. We used Miracl and Mathematica to find these prime factors. An $R$ stands for the remaining, not factored, part.

$$J_{239} - 1 = 2^{120} \cdot 3 \cdot 127 \cdot 239 \cdot 20231 \cdot 131071 \cdot 340337 \cdot 131105292137 \cdot 62983048367 \cdot R,$$

$$J_{757} - 1 = 2^{579} \cdot 5 \cdot 13 \cdot 29 \cdot 37 \cdot 109 \cdot 113 \cdot 757 \cdot 1429 \cdot 14449 \cdot 246241 \cdot 279073 \cdot 40388473189 \cdot 118750098349 \cdot 45637643105362633947353320957 \cdot 30483275619586522928407891468769 \cdot R,$$

$$J_{859} - 1 = 2^{430} \cdot 3^2 \cdot 67 \cdot 683 \cdot 859 \cdot 2003 \cdot 2731 \cdot 20857 \cdot 22366891 \cdot 6156182033 \cdot 10425285443 \cdot 15500487753323 \cdot 8340357737139637289786276330761 \cdot 1850748462483195301322746918526344689 \cdot R,$$

$$J_{1723} - 1 = 2^{862} \cdot 3^2 \cdot 43 \cdot 83 \cdot 739 \cdot 1723 \cdot 5419 \cdot 165313 \cdot 10388827 \cdot 8831418697 \cdot 20958338017 \cdot 78390786811 \cdot 13194317913029593 \cdot 84413238703660609 \cdot 4336790831080504259 \cdot R.$$
Chapter 6

Elliptic curves and pseudorandom sequences

Summary In this chapter we investigate an application of elliptic curves. In (Gong, Berson, and Stinson 2000) a pseudorandom sequence is obtained from an elliptic curve. They considered as elliptic curve $\mathcal{E}$ a supersingular elliptic curve defined over $\mathbb{F}_2$ given explicitly by a Weierstrass equation. They considered such a curve over an extension field $\mathbb{F}_{2^n}$ and took the sequence of points $P, 2P, 3P, \ldots, (n-1)P$, where $P$ is some point of the curve $\mathcal{E}$ defined over $\mathbb{F}_{2^n}$ and $N$ is its order. Finally, they applied the trace map from $\mathbb{F}_{2^n}$ to $\mathbb{F}_2$ to (some of) the coordinates of the points to obtain a sequence in $\mathbb{F}_2$. We investigate their construction in more detail and generalize it to non-supersingular elliptic curves and other characteristics. In particular we will calculate the autocorrelation of such sequences, which was not done in (Gong, Berson, and Stinson 2000)

6.1 Some facts about elliptic curves

In this section we describe the basic theory of elliptic curves. Most of these statements can be found in the literature, in which case we will give a reference.

An elliptic curve defined over $\mathbb{F}_q$ is a pair $(\mathcal{E}, \mathcal{O})$, where $\mathcal{E}$ is an algebraic curve of genus 1 and $\mathcal{O}$ is some point of $\mathcal{E}$ defined over $\mathbb{F}_q$. It is always possible to find a plane model of $\mathcal{E}$ such that it is given by an equation of the form

\[ Y^2 + a_1 XY + a_3 Y = X^3 + a_2 X^2 + a_4 X + a_6 \]

and moreover the point $\mathcal{O}$ is the point $[0 : 1 : 0]$ at infinity. Such an equation is called a Weierstrass equation. As a reference see (Silverman 1986).

An elliptic curve can be made into an abelian group $(\mathcal{E}, \oplus)$ with identity $\mathcal{O}$. Before defining the addition $\oplus$, we introduce some constants. Define
\[ b_2 = a_1^2 + 4a_2, \]
\[ b_4 = 2a_4 + a_1a_3, \]
\[ b_6 = a_5^2 + 4a_6, \]
\[ b_8 = a_1^3a_6 + 4a_2a_6 - a_1a_5a_4 + a_2a_4^2 - a_1^2. \]

Write \( P_1 = (x_1, y_1) \) and \( P_2 = (x_2, y_2) \). We start defining addition by

\[ P_1 \oplus P_2 = \mathcal{O} \text{ if } x_1 = x_2 \text{ and } y_2 = -y_1 - a_1x - a_3. \]

In this case \( P_1 = \ominus P_2 \). In the other cases \( P_1 \oplus P_2 \) is an affine point \((x_3, y_3)\) defined by

\[
x_3 = \begin{cases} 
\frac{y_2-y_1^2}{x_2-x_1} + a_1 \frac{y_2-y_1}{x_2-x_1} - a_2 - x_1 - x_2 & \text{if } x_1 \neq x_2, \\
\frac{x^4-4a_2x^2+2a_4x-a_6}{4x^3+2a_2x^2+2a_4x+a_6} & \text{if } x_1 = x_2
\end{cases}
\]

and

\[
y_3 = \begin{cases} 
-(\frac{y_2-y_1}{x_2-x_1} + a_1)x_3 - \frac{y_1x_2-y_2x_1}{x_2-x_1} - a_3 & \text{if } x_1 \neq x_2, \\
-(\frac{3x_1^2+2a_2x_1+a_4-a_1y_1}{2y_1+a_1x_1+a_3} + a_1)x_3 - \frac{-x_1^3+a_4x_1^2+2a_6-a_3y_1}{2y_1+a_1x_1+a_3} - a_3 & \text{if } x_1 = x_2.
\end{cases}
\]

These formulas were taken from p.58-59 of (Silverman 1986).

Now that we have a group structure on an elliptic curve, we would like to describe its group structure. The following proposition does this for the case that the elliptic curve is defined over a finite field \( \mathbb{F}_q \) (see p.145 of (Silverman 1986)).

**Proposition 6.1.1** Let \( \mathcal{E} \) be an elliptic curve defined over the finite field \( \mathbb{F}_q \). There exist numbers \( k \) and \( l \) such that as abelian groups

\[ \mathcal{E}(\mathbb{F}_q) \cong \mathbb{Z}/k\mathbb{Z} \times \mathbb{Z}/l\mathbb{Z}. \]

Further we have \( k|(q-1) \).

Finally, we make some remarks about addition on supersingular elliptic curves. On supersingular elliptic curves the addition is more “compatible” with the trace map than on other elliptic curves. Before stating a more precise result we need a theorem which classifies all supersingular elliptic curves. As a reference see Theorem 4.1 of (Waterhouse 1969).

**Theorem 6.1.2** Let \( \mathcal{E} \) be an elliptic curve over the finite field \( \mathbb{F}_q = \mathbb{F}_{p^e} \). Denote by the \( t \) the number \(|E(\mathbb{F}_q)| - q - 1 \). The curve \( \mathcal{E} \) is supersingular if and only if one of the following holds

Case i) \( e \) is odd and \( t = 0 \),

Case ii) \( e \) is odd, \( p = 2 \) and \( t = \pm \sqrt{2q} \),
6.1 Some facts about elliptic curves

Case iii) $e$ is odd, $p = 3$ and $t = \pm \sqrt{3q}$,
Case iv) $e$ is even, $p \not\equiv 1 \pmod{4}$ and $t = 0$,
Case v) $e$ is even, $p \not\equiv 1 \pmod{3}$ and $t = \pm \sqrt{q}$,
Case vi) $e$ is even and $t = \pm 2\sqrt{q}$.

Denote by $[m]$ the multiplication by $m$ map on an elliptic curve. Further denote by $\varphi_q$ the Frobenius map on an elliptic curve defined by $\varphi_q(x, y) = (x^q, y^q)$.

**Proposition 6.1.3** Let $E$ be a supersingular elliptic curve with trace $t$ defined over the finite field $\mathbb{F}_q$. Suppose first that $\mathbb{F}_q$ is an odd extension of the prime field $\mathbb{F}_p$. Then:
if $t = 0$, then $[q] = -\varphi_q^2$,
if $t^2 = 2q$, then $[q^2] = -\varphi_q^4$,
if $t^2 = 3q$, then $[q^3] = -\varphi_q^6$.
Now suppose that $\mathbb{F}_q$ is an even extension of the prime field $\mathbb{F}_p$. Then:
if $t = 0$, then $[q] = -\varphi_q^2$,
if $t = \pm \sqrt{q}$, then $[q\sqrt{q}] = \mp \varphi_q^3$,
if $t = \pm 2\sqrt{q}$, then $[\sqrt{q}] = \pm \varphi_q$.

**Proof:** All these statements are based on the fact that if an elliptic curve (supersingular or not) defined over a finite field $\mathbb{F}_q$ has trace $t$, then

$$\varphi_q^2 \cdot [t] \circ \varphi_q + [q] = 0.$$ 

So we know for example that $t = 0$ implies that $[q] = -\varphi_q^2$. The other cases follow similarly. Some care should be taken in the case that $t = \pm 2\sqrt{q}$. In this case we know that $\varphi_q^2 = 2\sqrt{q}\varphi_q + q = 0$ and hence $(\varphi_q \pm \sqrt{q})^2 = 0$. However, since the endomorphism ring of a supersingular elliptic curve can be embedded in a (definite) quaternion algebra (see Chapter V, Theorem 3.1 of (Silverman 1986)), it does not have zero divisors. Hence we find $\varphi_q = \pm \sqrt{q}$. □

For supersingular elliptic curves Proposition 6.1.1 can be refined. We quote the following proposition (see p.82, Lemma 4.8 of (Schoof 1985)).

**Proposition 6.1.4** Let $t \in \mathbb{Z}$. Denote by $I(t)$ the set of elliptic curves with trace $t$, up to isogeny. Denote by $p$ the characteristic and by $h(\cdot)$ the ordinary class number function.

If $t^2 = q, 2q$ or $3q$, then for every curve $E$ in $I(t)$ the group $E(\mathbb{F}_q)$ is cyclic.

If $t = \pm 2\sqrt{q}$, then for every curve $E$ in $I(t)$ we have

$$E(\mathbb{F}_q) \cong \mathbb{Z}/(\sqrt{q} + 1)\mathbb{Z} \times \mathbb{Z}/(\sqrt{q} - 1)\mathbb{Z}.$$ 

If $q \not\equiv -1 \pmod{4}$, then for every curve $E$ in $I(0)$ the group $E(\mathbb{F}_q)$ is cyclic.

If $q \equiv -1 \pmod{4}$, then for $h(-4p)$ curves $E$ in $I(0)$ the group $E(\mathbb{F}_q)$ is cyclic, while for the other $h(-p)$ curves in $I(0)$ we have
\[ E(F_q) \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/((q+1)/2)\mathbb{Z} \]

6.2 Pseudorandom sequences from elliptic curves

In this section we will give a construction of pseudorandom sequences and investigate it. As a result we will obtain sequences with good balance and autocorrelation. We start by giving the basic construction and, before that, some definitions.

**Definition 6.2.1** Let \( C \) be an algebraic curve defined over a finite field \( F_{q^e} \) and with function field \( \mathcal{F}(C) \). Let \( f \in \mathcal{F}(C) \) be a function on \( C \). Define \( C(f, F_{q^e}) \) to be the set of points in \( \mathcal{C}(F_{q^e}) \) for which there exists a \( g \in \mathcal{F}(C) \) such that the function \( f - g^q + g \) is defined at \( P \). From now on we will assume that for every point \( P \) not in \( C(f, F_{q^e}) \) there exists a function \( g \) defined over \( F_{q^e} \) such that \( v_P(f - g^q + g) \) is less than zero and not divisible by \( p \). Note that if \( q = p \), this condition is always fulfilled (see p. 114 of (Stichtenoth 1993)). Further note that the quantity \( Tr_{F_{q^e}/F_q}((f - g^q + g)(P)) \) does not depend on which \( g \) we choose, as long as \( f - g^q + g \) is defined at \( P \). This is why we will usually write \( Tr_{F_{q^e}/F_q}(f(P)) \) in this situation even if \( f \) itself is not defined in \( P \). Of course if \( f \) itself is defined at \( P \) we can choose \( g = 0 \).

We define the multiset

\[
S(f) = \{ Tr_{F_{q^e}/F_q}(f(P)) \mid P \in C(f, F_{q^e}) \},
\]

where \( Tr_{F_{q^e}/F_q} \) is the trace map from \( F_{q^e} \) to \( F_q \) defined by

\[
Tr_{F_{q^e}/F_q}(a) = \sum_{i=0}^{e-1} a^i.
\]

Finally, define for \( a \in F_q \) the natural number \( T_f(a) \) to be the multiplicity of \( a \) in the multiset \( S(f) \).

We want to calculate the numbers \( T_f(a) \). The following theorem essentially gives these numbers. First we state a lemma. It was first remarked by Serre (unpublished) and later by Lachaud (see (Lachaud 1987)).

**Lemma 6.2.2** Let \( C_1 \) and \( C_2 \) be algebraic curves defined over \( F_q \). Suppose that an algebraic map \( \phi \) exists from \( C_2 \) to \( C_1 \) defined over \( F_q \) as well. Then the set of Frobenius eigenvalues of \( C_1 \) is a subset of the set of Frobenius eigenvalues of \( C_2 \).

A consequence of this lemma is a generalization of the Serre bound, as stated a in (Lachaud 1991).

**Proposition 6.2.3** Let \( C_1 \) (respectively \( C_2 \)) be algebraic curves defined over \( F_q \) of genus \( g_1 \) (respectively \( g_2 \)) and suppose that there exist an algebraic map \( \phi \) from \( C_2 \) to \( C_1 \) defined over \( F_q \) as well. We have
\[ |\#C_2(F_q) - \#C_1(F_q)| \leq (g_2 - g_1)|2\sqrt{q}|, \]

where \(|\cdot|\) denotes the entier or floor function.

We apply the above proposition to obtain an upper bound in the following theorem. It is in essence already contained in p. 18 of (Voss 1993), but we still give the proof for the reader’s convenience. The case of genus zero can also be found in Chapter VIII, Section 2 of (Stichtenoth 1993).

**Theorem 6.2.4** Let \( C_1 \) be an algebraic curve of genus \( g_1 \) defined over \( \mathbb{F}_q \) with function field \( F_1 \). Let \( f \in F_1 \) and suppose that \( \Theta^2 - \Theta - f \in F_1[\Theta] \) is an absolutely irreducible polynomial. Assume that for every point \( P \) not in \( C_1(f, \mathbb{F}_q) \) there exists a function \( g \) defined over \( \mathbb{F}_q \) such that \( v_P(f - g^2 + g) \) is less than zero and not divisible by \( p \). Write \( m_P = -v_P(f - g^2 + g) \) in this case and \( m_P = -1 \) if \( P \in C_1(f, \mathbb{F}_q) \).

Denote by \( F_2 \) the function field \( F_1(\Theta) \) with \( \Theta^2 - \Theta = f - \alpha \), where \( \alpha \in \mathbb{F}_q^* \) is such that \( Tr_{\mathbb{F}_q^*|F_q}(\alpha) = a \). Denote the corresponding curve by \( C_2 \). We have

\[
T_f(a) = \frac{\#C_2(F_{q^t}) - \#C_1(F_{q^t}) + \#C_1(f, F_{q^t})}{q},
\]

The genus \( g_2 \) of \( C_2 \) does not depend on \( a \). Further we have

\[
\left| T_f(a) - \frac{\#C_1(f, F_{q^t})}{q} \right| \leq \frac{(g_2 - g_1)}{q}|2\sqrt{q}|.
\]

Finally

\[
g_2 = gg_1 + \frac{q - 1}{2} \left( -2 + \sum_{P} (m_P + 1) \right),
\]

where the sum ranges over all \( P \in C_1(f, \mathbb{F}_q) \).

**Proof:** A point \( P \in C(f, \mathbb{F}_q^t) \) satisfies \( Tr_{\mathbb{F}_q|\mathbb{F}_q}(f(P)) = 0 \) if and only if there exist a \( \Theta \in F_{q^t} \) such that \( \Theta^2 - \Theta = f(P) \). This follows from Hilbert’s Theorem 90 (see for example Chapter VI, Theorem 6.1 of (Lang 1993)). Since \( Tr_{\mathbb{F}_q|\mathbb{F}_q}(f(P)) = a \) if and only if \( Tr_{\mathbb{F}_q|\mathbb{F}_q}(f(P) - \alpha) = 0 \), where \( Tr_{\mathbb{F}_q|\mathbb{F}_q}(\alpha) = a \), this implies that \( Tr_{\mathbb{F}_q|\mathbb{F}_q}(f(P)) = a \) if and only if there exists a \( \Theta \in \mathbb{F}_q^t \) such that \( \Theta^2 - \Theta = f(P) - \alpha \). If this equation has a solution \( \Theta \in \mathbb{F}_q^t \), all \( q \) solutions are defined over \( \mathbb{F}_q^t \), since they are given by \( \Theta + a \), where \( a \in \mathbb{F}_q \). Further any point of \( C_1(F_{q^t}) \) not in \( C_1(f, F_{q^t}) \) corresponds to exactly one point of \( C_2(F_{q^t}) \). Hence we see that the number \( \#C_2(F_{q^t}) - \#C_1(F_{q^t}) + \#C_1(f, F_{q^t}) \) is \( q \) times the number \( T_f(a) \).

The function field \( F(C_2) \) depends on \( a \). However the function field \( \mathcal{G}(C_2) \) obtained from \( F(C_2) \) by extending the field of constants to \( \mathbb{F}_{q^t} \) does not. Hence the number \( g_2 \) does not depend on \( a \). The explicit expression for \( g_2 \) follows directly from Proposition 1.3.2 The rest of the theorem follows from Hasse-Weil’s theorem, Lemma 6.2.2 and Proposition 6.2.3. \( \square \)
Note that \( g_2 \) and \( g_1 \) do not depend on \( e \) in the sense that if we enlarge the field of definition \( \mathbb{F}_{q^e} \) to, say, \( \mathbb{F}_{q^f} \), \( g_2 \) and \( g_1 \) do not change. We will now go to the study of sequences.

**Definition 6.2.5** Let \( C \) be an algebraic curve defined over \( \mathbb{F}_{q^e} \) and \( f \in \mathcal{F}(C) \). Enumerate the points of \( \mathcal{C}(f, \mathbb{F}_{q^e}) \) as \( P_1, \ldots, P_N \). We then define the sequence

\[
S_f = \text{Tr}_{\mathbb{F}_{q^e}:\mathbb{F}}(f(P_1)^e), \text{Tr}_{\mathbb{F}_{q^e}:\mathbb{F}}(f(P_2)^e), \ldots, \text{Tr}_{\mathbb{F}_{q^e}:\mathbb{F}}(f(P_N)^e).
\]

Note that the number of times a symbol \( a \in \mathbb{F}_q \) occurs in this sequence is exactly \( T_f(a) \). We now consider the balance of the sequences we have constructed.

**Definition 6.2.6** Let \( S = s(0), s(1), \ldots, s(N) \) be a sequence of elements of \( \mathbb{F}_q \). Let \( \alpha \in \mathbb{F}_q^\times \). We define the balance with respect to \( \alpha \) in the following way.

\[
B_S(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \zeta_p(\text{Tr}_{\mathbb{F}_{q^e}:\mathbb{F}}(\alpha s(i))^e),
\]

with \( \zeta_p = e^{2\pi i/p} \). Further we define the balance to be

\[
B_S = \max_{\alpha \in \mathbb{F}_q^\times} |B_S(\alpha)|.
\]

If a sequence contains all elements of \( \mathbb{F}_q \) the same number of times the balance is 0, but if a sequence contains only one element the balance is 1. To obtain an upperbound for the balance of the above sequences, one could use Theorem 6.2.4. However, it is possible to obtain better results using some known results about exponential sums. We quote a theorem (see for example (Voloch and Walker 2000)).

**Theorem 6.2.7** Let \( C \) be an algebraic curve defined over \( \mathbb{F}_{q^e} \) of genus \( g_1 \) and \( \mathcal{F} \) be its function field. Choose \( f \in \mathcal{F} \) and denote by \( \mathcal{C}(f, \mathbb{F}_{q^e}) \) the set of \( \mathbb{F}_{q^e} \)-rational points \( P \) for which \( g \in \mathcal{F} \) exists (depending on \( P \) ) such that the function \( f - g^2 + g \) is defined at \( P \). Suppose that the polynomial \( \Theta - \Theta - f \in \mathcal{F}(\Theta) \) is absolutely irreducible and denote its splitting field by \( \mathcal{G} \). Write \( D \) for the corresponding curve and suppose that it has genus \( h \). We have

\[
\left| \sum_{P \in \mathcal{C}(f, \mathbb{F}_{q^e})} \zeta_p(\text{Tr}_{\mathbb{F}_{q^e}:\mathbb{F}}((f - g^2 + g)(P))^e) \right| \leq \frac{2(h - g_1)}{p - 1} \sqrt{q^e}.
\]

**Corollary 6.2.8** Let the notation be the same as above. Suppose that for every point \( P \) of \( C \) there exists function \( g \) such that either the characteristic does not divide \( v_P(f - g^2 + g) \) or \( v_P(f - g^2 + g) \geq 0 \). We have

\[
B_{S_f} \leq \frac{1}{\#\mathcal{C}(f, \mathbb{F}_{q^e})} \left( 2g_1 - 2 + \sum_{P} (1 - v_P(f)) \right) \sqrt{q^e},
\]
6.2 Pseudorandom sequences from elliptic curves

where the sum ranges over the points $P$ such that $f$ has a pole at $P$, but $P \notin \mathcal{C}(f; \mathbb{F}_q^*)$.

**Proof:** Note that under the above assumptions the set of $\mathbb{F}_q^*$-rational points such that there exists a $g$ such that $f - g^q + g$ is defined at $P$ is the same as the set of $\mathbb{F}_q^*$-rational points such that there exists a $g$ such that $f - g^q + g$ is defined at $P$. Hence we can use the same notations $\mathcal{C}(f; \mathbb{F}_q^*)$ for both of them. Applying the previous theorem gives an upper bound for $B_{S_f}$. From Proposition 1.3.2 we see that $h = pg_1 + (p-1)(-2 + \sum P (m_P + 1)) / 2$. Moreover we have $m_R + 1 = 0$ if $f$ does not have a pole at $P$ or $P \in \mathcal{C}(f; \mathbb{F}_q^*)$. For the other points $P$ we have $m_P \leq -v_P(f)$. This gives an upper bound for $h$. Putting the two upper bounds together gives the desired result. \qed

It can happen that for small $\epsilon$ the balance is not good, but by enlarging the field of definition, we can approach perfect balance as close as we want.

The problem of Definition 6.2.5 is that we have to choose an ordering of the points in the set $\mathcal{C}(f; \mathbb{F}_q^*)$. If we are unlucky in our choice, we could end up with a sequence that looks very regular. The problem is that for general curves, as far as we know, no natural ordering of the points is available. For elliptic curves the situation can be different. Suppose for example that $\mathcal{E}$ is an elliptic curve and that the group $E(\mathbb{F}_q^*)$ is cyclic of order $N$ with generator $P$. In this case the ordering $O, P, [2]P, [3]P, \ldots, [N-1]P$ is very natural (we denoted the zero of the group by $O$).

**Definition 6.2.9** Let $\mathcal{E}$ be an elliptic curve defined over the finite field $\mathbb{F}_q^*$. Let $P$ be a point defined over $\mathbb{F}_q^*$ of order $N$ and let $f$ a function defined on all points $P; [2]P; \ldots, [N-1]P$. We define the *basic sequence* corresponding to $f$ and $P$ to be the sequence

$$S_{f, P} = Tr_{\mathbb{F}_q^*[\mathbb{F}_q]}(f(O)), Tr_{\mathbb{F}_q^*[\mathbb{F}_q]}(f(P)), \ldots, Tr_{\mathbb{F}_q^*[\mathbb{F}_q]}(f([N-1]P)).$$

If $O \notin \mathcal{E}(f; \mathbb{F}_q^*)$, we use the convention that $Tr_{\mathbb{F}_q^*[\mathbb{F}_q]}(f(O)) = 0$. The reason we insist to include this term in our pseudorandom sequence will become clear when we analyze the autocorrelation.

Note that in case $P$ is a generator for the group $\mathcal{E}(\mathbb{F}_q^*)$, the balance of this sequence can be estimated using Corollary 6.2.8. We will be more specific about the balance later. To measure how random this sequence is, we introduce the following quantity.

**Definition 6.2.10** Let $s(1), s(2), \ldots, s(N)$ be a pseudorandom sequence $S$ of period $N$ defined over the finite field $\mathbb{F}_q^*$. Write $p$ for the characteristic. We define the *autocorrelations* of a sequence as follows:

$$C_S(d, \alpha, \beta) = \frac{1}{N} \sum_{i=1}^{N} \zeta_p^{Tr_{\mathbb{F}_q^*[\mathbb{F}_q]}(\alpha s(i+d) - \beta s(i))},$$

with $0 \leq d < n$, $\alpha, \beta \in \mathbb{F}_q^*$ and $\zeta_p = \exp(2\pi i / p)$.
For binary sequences this definition amounts to

\[ C_S(d) = \frac{1}{N} \sum_{i=1}^{N} (-1)^{s_i + d^* s_i}, \]

which is the usual definition of the autocorrelation (see for example Chapter 5, Section 4 of (Menezes, van Oorschot, and Vanstone 1997)). We will estimate these autocorrelations below. First we give more specific results about the balance than before.

**Proposition 6.2.11** Let \( E \) be an elliptic curve defined over \( \mathbb{F}_{q^r} \) given by the Weierstrass equation

\[ Y^2 + a_1 XY + a_3 Y = X^3 + a_2 X^2 + a_4 X + a_6, \]

Suppose that \( E(\mathbb{F}_{q^r}) \) is a cyclic group of order \( N \) with generator \( P \). Let \( f \) be a non-constant polynomial in the coordinate functions \( x \) and \( y \). We assume without loss of generality that \( \deg_y(f) \leq 1 \). Denote by \( \text{wdeg} \) the weighted degree defined by \( \text{wdeg}(x) = 2 \) and \( \text{wdeg}(y) = 3 \). Denote by \( S_{f,P} \) the \( q \)-ary basic sequence corresponding to \( f \) and \( P \). Denote by \( \mathcal{C} \) the curve given by the equations

\[ Y^2 + a_1 XY + a_3 Y = X^3 + a_2 X^2 + a_4 X + a_6, \]

and

\[ \Theta^p - \Theta = f(X,Y). \]

If the characteristic does not divide \( \text{wdeg}(f) \), then this curve is absolutely irreducible and has genus \( 1 + \frac{p-1}{2}(1 + \text{wdeg}(f)) \). In this case we have

\[ B_{S_f,P} \leq \frac{1}{\#E(\mathbb{F}_{q^r})} \left( (1 + \text{wdeg}(f))\sqrt{q^r + 1} \right). \]

**Proof:** Most of this follows immediately from Corollary 6.2.8. We have a natural covering \( \phi \) from \( \mathcal{C} \) to \( E \) defined (on the affine piece) by

\[ \phi(x, y, \theta) = (x, y). \]

We will use Proposition 1.3.2 to analyze this covering. For all affine points \( P \) of \( E \) we see that \( m_P = -1 \), since for such points we have

\[ v_P(f(x, y)) \geq 0. \]

For the point \( \mathcal{O} = [0 : 1 : 0] \), we have

\[ v_\mathcal{O}(f(x, y)) = -\text{wdeg}(f), \]

since from the assumption \( \deg_y(f(X,Y)) \leq 1 \) we see that exactly one monomial \( x^i y^j \) occurring in \( f \) satisfies \( 2i + 3j = \text{wdeg}(f) \). Hence \( m_{\mathcal{O}} = \text{wdeg}(f) \). We find the
6.2 Pseudorandom sequences from elliptic curves

desired formula for the genus directly from this. Further, we see that \( \mathcal{O} \not\in \mathcal{E}(f, \mathbb{F}_{q^r}) \), but that all other points in \( \mathcal{E}(\mathbb{F}_{q^r}) \) are contained in this set. This explains the term 1 in the upper bound for the balance.

The question remains about what happens if the characteristic \( p \) does divide \( \text{wdeg}(f) \). In this case several things can happen. The curve \( \mathcal{C} \) could be reducible over \( \mathbb{F}_{q^r} \). In that case a function \( g \in \mathcal{F}(\mathcal{E}) \) exists such that \( f = g^p - g \). In this case \( S_{f, p} \) is the all zero sequence and hence \( B_{S_{f, p}} = 1 \). Further \( \mathcal{C} \) could be reducible, but only over \( \mathbb{F}_{q^{r'}} \). In that case we have a constant field extension. This means that the sequence \( S_{f, p} \) is a multiple of the all one sequence (apart from the term corresponding with \( \mathcal{O} \)). If on the other hand \( \mathcal{C} \) is absolutely irreducible, the above expression for the genus becomes a strict upper bound, since the value of \( m_{\mathcal{O}} \) will become strictly less than \( \text{wdeg}(f) \). Hence the upper bound for \( B_{S_{f, p}} \) will be strict as well. We will give some examples later.

We will now turn our attention to the autocorrelation. If \( P \) is a point of an elliptic curve, then we write \((X, Y) \oplus P\) as shorthand for the polynomials occurring in the addition formula of \( \mathcal{E} \) mentioned in the previous section.

**Theorem 6.2.12** Let \( \mathcal{E} \) be a cyclic elliptic curve defined over the finite field \( \mathbb{F}_{q^s} \) of characteristic \( p \), given by the Weierstrass equation

\[
Y^2 + a_1 XY + a_3 Y = X^3 + a_2 X^2 + a_4 X + a_6,
\]

Let \( P \) be a generator of the group \( \mathcal{E}(\mathbb{F}_{q^r}) \). Let \( f \) be a polynomial in the coordinate functions \( x \) and \( y \) with the property \( \text{deg}_y(f) \leq 1 \). Fix \( d \) with \( 1 \leq d < \#\mathcal{E}(\mathbb{F}_{q^r}) \) and \( \alpha, \beta \in \mathbb{F}_q^* \). Denote by \( \mathcal{D} \) the curve (depending also on \( d, \alpha \) and \( \beta \)) given by the equations

\[
Y^2 + a_1 XY + a_3 Y = X^3 + a_2 X^2 + a_4 X + a_6
\]

and

\[
\Theta^p - \Theta = \alpha f((X, Y) \oplus [d]P) - \beta f(X, Y),
\]

Suppose that \( \mathcal{D} \) is absolutely irreducible, for which a sufficient condition is that \( p \) does not divide \( \text{wdeg}(f) \). Its genus \( g \) satisfies

\[
g \leq 1 + (p - 1)(1 + \text{wdeg}(f)),
\]

the inequality being sharp if \( p \nmid \text{wdeg}(f) \). Further we have

\[
|C_{S_{f, p}}(d, \alpha, \beta)| \leq \frac{1}{\#\mathcal{E}(\mathbb{F}_{q^s})} \left( 2 + 2 \frac{g - 1}{p - 1} \sqrt{q^s} \right).
\]

**Proof:** If \( p \nmid \text{wdeg}(f) \), we have \( m_{\mathcal{O}} = m_{[-d]P} = \text{wdeg}(f) \). Hence from Proposition 1.3.2 we see that \( \mathcal{D} \) is absolutely irreducible in that case. Suppose that \( \mathcal{D} \) is absolutely irreducible. The upper bound for the genus is obtained by remarking that for all points \( Q \) of \( \mathcal{E} \) except \( \mathcal{O} \) and \( [-d]P \) we have \( m_Q = -1 \). Further \( m_{\mathcal{O}} \)
and $m_{[-d]P}$ are bounded from above by $\text{wdeg}(f)$. To obtain the upperbound for the autocorrelation, note that

$$\#\mathcal{E}(\mathbb{F}_{q^e}) C_{S, h}(d, \alpha, \beta) = \alpha \left[ \text{Tr}_{\mathbb{F}_{q^e} / \mathbb{F}_q}(\alpha f([d]P)) \right] + \beta \left[ \text{Tr}_{\mathbb{F}_{q^e} / \mathbb{F}_q}(\beta f([d]P)) \right] + \#\mathcal{E}(h, \mathbb{F}_{q^e}) B_{S, h},$$

with $h = \alpha f((x, y) \oplus [d]P) - \beta f(x, y)$. Hence the upperbound follows from Corollary 6.2.8. \hfill \Box

**Corollary 6.2.13** Let the notation be the same as in the previous theorem. Choose $\alpha = \beta = 1$ and suppose that the corresponding curve $D$ is absolutely irreducible and has genus $g$. If

$$\#\mathcal{E}(\mathbb{F}_{q^e}) > 2 + \frac{q + 1}{p - 1},$$

then the basic sequence $S_{f, P}$ has full period $\#\mathcal{E}(\mathbb{F}_{q^e})$.

**Proof:** Assume that the basic sequence has period $k$ and write $l = \#\mathcal{E}(\mathbb{F}_{q^e})/k$. Then we have

$$C_{S, h}(k, 1, 1) = 1.$$ 

This gives a contradiction with the upper bound from the previous theorem. \hfill \Box

Until now we have always assumed that the group $\mathcal{E}(\mathbb{F}_{q^e})$ is cyclic. This condition is not always fulfilled and even if it is, one might want to work in a subgroup of prime order. It is possible in some cases to generalize the above results to cyclic subgroups of a certain kind. We will now indicate an approach on how to do this. The kind of cyclic subgroups for which we can generalize the above results, will have the form $[D]\mathcal{E}(\mathbb{F}_{q^e})$, where $D$ is multiple of the number $k(\mathcal{E})$ which we will define now.

**Definition 6.2.14** Let $\mathcal{E}$ be an elliptic curve defined over $\mathbb{F}_{q^e}$. Define the natural number $k(\mathcal{E}, f)$ to be the largest integer $n$ such that $\mathcal{E}(\mathbb{F}_q)[n] \subset \mathcal{E}(\mathbb{F}_{q^e})$ and $n$ is relatively prime to the characteristic. We will write $k(\mathcal{E})$ as shorthand notation for $k(\mathcal{E}, 1)$.

Trivially, the group $\mathcal{E}(\mathbb{F}_{q^e})$ of an elliptic curve defined over $\mathbb{F}_{q^e}$, is cyclic if and only if $k(\mathcal{E}) = 1$. From Proposition 6.1.1 we see that $k(\mathcal{E})$ divides $q^e - 1$. If $D$ is a multiple of $k(\mathcal{E})$, then obviously $[D]\mathcal{E}(\mathbb{F}_{q^e})$ is cyclic. Moreover we can assume that $D$ divides $\#\mathcal{E}(\mathbb{F}_{q^e})/k(\mathcal{E})$. Using the addition formula on $\mathcal{E}$ one can express $[D](x, y)$ as rational functions in $x$ and $y$. We now state the following lemma, which is the key to the generalization of our results about the weights.

**Lemma 6.2.15** Let $\mathcal{E}$ be an elliptic curve defined over the finite field $\mathbb{F}_{q^e}$ with $q = p^m$. Let $f \in \mathcal{F}(\mathcal{E})$. Suppose that $D \in \mathbb{N}$ is such that $k(\mathcal{E}) \mid D \mid \#\mathcal{E}(\mathbb{F}_{q^e})/k(\mathcal{E})$ and write $D = Dp^m$, with $\gcd(p, D) = 1$. Denote by $P$ a generator of $[D]\mathcal{E}(\mathbb{F}_{q^e})$. Finally denote by $S$ the sequence $\{\text{Tr}_{\mathbb{F}_{q^e} / \mathbb{F}_q}(f(Q))\}$, where $Q \in \mathcal{E}(f, \mathbb{F}_{q^e}) \cap < P >$. We have
6.2 Pseudorandom sequences from elliptic curves

\[ BS(\alpha) = BS_{f \restriction \alpha}(\alpha). \]

Suppose that \( \Theta^q - \Theta - f \circ D \in \mathcal{F}(E) \) is absolutely irreducible and denote the curve corresponding with its splitting field by \( C \). Further write \( g \) for its genus. We have

\[ BS \leq \frac{1}{\# \mathcal{E}(f \circ D, \overline{\mathbb{F}_{q^r}})} \frac{2(g - 1)}{p - 1} \sqrt{q^r} \]

and

\[ g \leq 1 + \frac{(p - 1)\bar{D}^2p^\alpha}{2} \sum_Q (1 - p^n v_Q(f)), \]

where the sum is taken over all points \( Q \not\in \mathcal{E}(f, \overline{\mathbb{F}_{q^r}}) \) for which \( f \) has a pole. The inequality for the genus is sharp if \( n = 0 \) and the occurring \( v_Q(f) \) are relatively prime to \( p \).

**Proof:** We have \([D] \mathcal{E}(f \circ D, \overline{\mathbb{F}_{q^r}}) = \mathcal{E}(f, \overline{\mathbb{F}_{q^r}}) \cap \{ P < P \} \). Further, for each point \( R \) in the latter set, there exist \( Dk(\mathcal{E}) \) points \( Q \) in the former set such that \([D]Q = R\). Hence

\[ \# \mathcal{E}(f \circ D, \overline{\mathbb{F}_{q^r}}) BS_{f \restriction \alpha}(\alpha) = Dk(\mathcal{E}) \#(\mathcal{E}(f, \overline{\mathbb{F}_{q^r}}) \cap \{ P < P \}) BS(\alpha). \]

The upper bound for \( BS \) now follows immediately from Corollary 6.2.8. The statements about the genus follow from Proposition 1.3.2. Note that for every pole \( R \) of \( f \) there exist \( p^n \bar{D}^2 \) poles \( Q \) of \( f \circ D \). Further for each of these \( Q \) we have \( v_Q(f \circ D) = p^n v_R(f) \).

\[ BS_{f \restriction \nu} \leq \frac{1}{\# < P >} \left( 1 + \frac{D}{k(\mathcal{E})} (1 + w deg(f)) \sqrt{q^r} \right). \]

**Proof:** The proof is similar to the case \( D = 1 \). We find

\[ (\# < P >) BS_{f \restriction \nu} = 1 + (\# < P > - 1) BS, \]

with \( S \) the sequence \( \{ f(Q) \} \), \( Q \in \mathcal{E}(f, \overline{\mathbb{F}_{q^r}}) \cap \{ P < P \} \). After applying the upper bound for \( BS \) in the previous lemma we find the desired result.\( \square \)
Corollary 6.2.17 Let the notation be the same as above. For \( D = k(\mathcal{E}) \) we find
\[
B_{S_{f,p}} \leq \frac{1}{\# < P>} \left( 1 + (1 + \text{wdeg}(f))\sqrt{q^p} \right).
\]
The results for the autocorrelation can be generalized in a similar manner. We state the results below.

Lemma 6.2.18 Let \( \mathcal{E} \) be an elliptic curve given by a Weierstrass equation defined over the field \( \mathbb{F}_{q^p} \) of characteristic \( p \). Let \( f \in \mathcal{F}(\mathcal{E}) \) and choose \( \alpha, \beta \in \mathbb{F}_{q^p}^* \). Further choose numbers \( D \) and \( d \) satisfying \( k(\mathcal{E}) \mid D \# \mathcal{E}(\mathbb{F}_{q^p}) / k(\mathcal{E}) \) and \( 1 \leq d < \# < P > \), with \( P \) a generator of the group \([D] \mathcal{E}(\mathbb{F}_{q^p})\). Write
\[
h(x, y) = \alpha f((x, y) \oplus [d]P) - \beta f(x, y).
\]
Suppose that the curve \( D \) defined as an extension of \( \mathcal{E} \) by the additional equation
\[
\Theta^p - \Theta = h \circ [D]
\]
is absolutely irreducible of genus, say, \( g \), for which a sufficient condition is that \( p \) does not divide \( D\text{wdeg}(f) \). Finally write \( D = D^p \), such that \( \text{gcd}(D, p) = 1 \). We have
\[
g \leq 1 + \bar{D}^2 p^n(p - 1) \sum_{Q}(1 + p^n v_Q(f)),
\]
where the sum is over all points \( Q \) such that \( f \) has a pole in \( Q \), but \( Q \notin \mathcal{E}(\mathbb{F}_{q^p}) \). We have equality if and only if \( p \) does not divide \( D\text{wdeg}(f) \). Further we have
\[
(\# < P >)C_{S_{f,p}}(d, \alpha, \beta) = c + \#(\mathcal{E}(\mathbb{F}_{q^p}) \cap < P >)B_{S_{h,\alpha,\beta}},
\]
with
\[
c = \zeta_p^{\tau_{\mathcal{E}}} \mathbb{F}_p(\alpha f([d]P)) + \zeta_p^{-\tau_{\mathcal{E}}} \mathbb{F}_p(\beta f([-d]P)).
\]

Proof: The proof is similar to that of Lemma 6.2.15 and left to the reader. \( \Box \)

Theorem 6.2.19 Let the notation be the same as in the above lemma. Let \( f \) be a polynomial in the two coordinate functions \( x \) and \( y \), such that \( \text{deg}_y(f) \leq 1 \). Suppose that the characteristic does not divide \( \text{wdeg}(f) \). We have
\[
|C_{S_{f,p}}(d, \alpha, \beta)| \leq \frac{1}{\# < P>} \left( 2 + 2 \frac{D}{k(\mathcal{E})} (1 + \text{wdeg}(f))\sqrt{q^p} \right).
\]

Proof: This follows immediately from the previous lemma and Corollary 6.2.8. \( \Box \)

We will now consider several examples over \( \mathbb{F}_2 \).
6.2 Pseudorandom sequences from elliptic curves

Example 6.2.20 In this example we will consider basic sequences obtained from the elliptic curve defined over \( \mathbb{F}_2 \) given by the Weierstrass equation

\[
Y^2 + Y = X^3 + X + 1.
\]

Over the field \( \mathbb{F}_2 \), this curve has Frobenius eigenvalues \((1 + i)^e\) and \((1 - i)^e\). Before proceeding we will calculate the numbers \( k(\mathcal{E}, e) \).

Lemma 6.2.21 For the curve mentioned above, we have

\[
k(\mathcal{E}, e)^2 = \begin{cases} 
1 & \text{if } e \not\equiv 0 \pmod{4}, \\
\#\mathcal{E}(\mathbb{F}_{q^e}) & \text{if } e \equiv 0 \pmod{4}.
\end{cases}
\]

Proof: This is clear immediately from Proposition 6.1.4. \( \square \)

This means that for \( e \equiv 0 \pmod{4} \) no non-trivial cyclic subgroup exists to which the above theory applies. Of course we can still calculate the quantities \( T_f(a) \) by using Theorem 6.2.4, but this has no use from the pseudorandom sequences point of view. From now on in this example we assume that \( e \not\equiv 0 \pmod{4} \). Denote by \( P \) a generator of the group \( \mathcal{E}(\mathbb{F}_{q^e}) \). We will examine the sequence \( S_{x, P} \) in some detail. First we investigate the balance.

First we need to calculate the genus \( g \) of the curve \( \mathcal{C} \) defined by the equations

\[
Y^2 + Y = X^3 + X + 1
\]

and

\[
\Theta^2 + \Theta = X.
\]

We have \( m_\mathcal{O} = 1 \), since \( v_\mathcal{O}(x + ((y/x)^2 + (y/x))) = v_\mathcal{O}(y/x + y/x^2 + 1/x + 1/x^2) = -1 \). Hence \( g = 2 \). Its Frobenius eigenvalues can be determined by looking at the number of points on this curve over two small fields. We find that they are given by \((1 \pm i)\), as predicted by Lemma 6.2.2 and \((\pm i\sqrt{2})\). We find

\[
T_\varepsilon(0) = \frac{\#\mathcal{E}(\mathbb{F}_{q^e}) - 1}{2} - \frac{(i\sqrt{2})^e + (-i\sqrt{2})^e}{2}
\]

and

\[
T_\varepsilon(1) = \frac{\#\mathcal{E}(\mathbb{F}_{q^e}) - 1}{2} + \frac{(i\sqrt{2})^e + (-i\sqrt{2})^e}{2},
\]

which can be reformulated by saying

\[
T_\varepsilon(0) = \begin{cases} 
\frac{\#\mathcal{E}(\mathbb{F}_{q^e}) - 1}{2} & \text{if } e \text{ is odd}, \\
\frac{\#\mathcal{E}(\mathbb{F}_{q^e}) - 1}{2} - (-1)^{e/2}2^{e/2+1} & \text{if } e \text{ is even},
\end{cases}
\]

and
\[ T_\epsilon(1) = \begin{cases} \frac{\# E(F_{q^\epsilon}) - 1}{2} & \text{if } \epsilon \text{ is odd,} \\ \frac{\# E(F_{q^\epsilon}) - 1}{2} + (-1)^{\epsilon/2} q^{\epsilon/2 + 1} & \text{if } \epsilon \text{ is even.} \end{cases} \]

For odd \( \epsilon \) these results are mentioned in (Gong, Berson, and Stinson 2000) as well. We obtain the additional result that for \( \epsilon \equiv 2 \pmod{4} \) the balance of the basic sequence with \( f = x \) is \( (2^{\epsilon/2 + 2} + 1)/(2^{\epsilon} + 1) \).

We will now investigate the autocorrelation of this sequence. This involves the calculation of the genus \( \tilde{g} \) of the curves \( D \) defined by

\[ Y^2 + Y = X^3 + X + 1 \]

and

\[ \Theta^2 + \Theta = x((x, y) \oplus [d] P) + x. \]

Using Theorem 6.2.12 we find \( \tilde{g} = 3 \). This implies

\[ |C_{S, \epsilon, \epsilon}(d)| \leq 2 + 4\sqrt{q}. \]

Hence Corollary 6.2.13 implies that the sequence \( S_{x, \epsilon} \) has full period if \( \epsilon \geq 4 \).

**Example 6.2.22** In this example we consider the elliptic curve \( E \) given by the equation

\[ Y^2 + XY = X^3 + 1, \]

defined over the field \( \mathbb{F}_2 \). First we will calculate the number \( k(E, \epsilon) \). This curve has four points defined over \( \mathbb{F}_2 \), hence all numbers \( \# E(F_{q^\epsilon}) \) are a multiple of four. Its Frobenius eigenvalues over \( \mathbb{F}_{q^\epsilon} \) are given by \(-1/2 \pm i\sqrt{7}/2\). To calculate the number \( k(E, \epsilon) \) is not as easy for this curve as it was in the last example. Using the property that if \( E(F_2)[n] \subseteq E(F_{2^\epsilon}) \), we have \( n^2 \mid \# E(F_{2^\epsilon}) \) and \( n\mid 2^\epsilon - 1 \), we can prove that \( k(E, \epsilon) = 1 \) for \( \epsilon = 1, 2, 3, 4, 5, 6, 7, 9 \). Moreover \( k(E, 8) = 3 \) and \( k(E, 10)|11 \).

Looking at more numerical data we conjecture that for this particular curve \( k(E, \epsilon) \) will be a small constant negligible to the total group order and that \( E(F_{2^\epsilon}) \) will be cyclic for infinitely many \( \epsilon \). For \( \epsilon \leq 10000 \) we found that for 5359 values of \( \epsilon \) the number \( \gcd(2^\epsilon - 1, \# E(F_{2^\epsilon})) \) is one and hence the elliptic curve \( E(F_{2^\epsilon}) \) is cyclic. We consider again the function \( x \). We find after some calculations

\[ T_\epsilon(0) = \frac{\# E(F_{q^\epsilon}) - 2}{2} \]

and

\[ T_\epsilon(1) = \frac{\# E(F_{q^\epsilon})}{2}. \]

Nice as this looks, the reason for this balance is very bad from the point of view of pseudorandomness. It is not hard to see that
\[ x([2](x, y)) = \frac{x + y^2}{x} + \frac{x + y}{x}. \]

Hence for all non-zero points \( Q \in [2] \mathcal{E}(\mathbb{F}_{2^r}) \) we have \( \text{Tr}_{\mathbb{F}_{2^r}/\mathbb{F}_2}(x(Q)) = 0 \) and for the other points \( Q \) we have \( \text{Tr}_{\mathbb{F}_{2^r}/\mathbb{F}_2}(x(Q)) = 1 \). We have \( \mathbb{C}_{\mathcal{E}, P}(2) = \#\mathcal{E}(\mathbb{F}_{2^r}) \) for any basic sequence. We will see if our methods yield this result as well. Assume for convenience that \( \mathcal{E}(\mathbb{F}_{2^r}) \) is cyclic with generator \( P \). We will investigate the curve \( \mathcal{D} \) given by the equations

\[ Y^2 + Y = X^3 + 1 \]

and

\[ \Theta^2 + \Theta = x((X, Y) \oplus [2]P) + X. \]

Write \( P = (\alpha, \beta) \) and \([2]P = (a, b)\). Note that

\[ a = \left( \frac{\alpha + \beta}{\alpha} \right)^2 + \frac{\alpha + \beta}{\alpha}. \]

Using the addition formula on \( \mathcal{E} \) we see that

\[ x((X, Y) \oplus [2]P) = \left( \frac{Y + b}{X + a} \right)^2 + \frac{Y + b}{X + a} + X + a. \]

Hence we can rewrite the second equation defining \( \mathcal{D} \) as

\[ \left( \Theta + \frac{\alpha + \beta}{\alpha} + \frac{Y + b}{X + a} \right)^2 + \Theta + \frac{\alpha + \beta}{\alpha} + \frac{Y + b}{X + a} = 0, \]

which shows that the curve \( \mathcal{D} \) is reducible over \( \mathbb{F}_{2^r} \). This explains in another way why the autocorrelation is not good. This does not mean that we cannot obtain good sequences for this curve. For other functions \( f \), for example \( f = y \), the balance and autocorrelation is good.

### 6.3 A family of sequences having small crosscorrelations

In this section we will describe several families of pseudorandom sequences having small crosscorrelations. Some of these results can be found in (Voloch and Walker 2000), although there sequences with entries in \( \mathbb{Z}/p^2\mathbb{Z} \) were considered, while we consider sequences with entries in a finite field. Further in this article the group \( \mathcal{E}(\mathbb{F}_{p^r}) \) is assumed to be cyclic, while we do not have this restriction. Finally the cyclic shifts of a sequence, resulting in the autocorrelation properties, were not considered there. First we define the crosscorrelation of two sequences.
Definition 6.3.1 Let $S = \{s(i)\}$ and $T = \{t(i)\}$ be two pseudorandom sequences defined over $\mathbb{F}_q$ having the same period $N$. Denote the characteristic of $\mathbb{F}_q$ by $p$ and let $\alpha, \beta \in \mathbb{F}_q^*$. We define the crosscorrelation by

$$C_{S,T}(d, \alpha, \beta) = \frac{1}{\# \mathbb{F}_q} \sum_{i=1}^{N} \zeta_p^{x \cdot (i+d) - \beta t(i)},$$

where $\zeta_p = \exp(2\pi i/p)$ and $0 \leq d < n$.

The problem is to find a family of sequences $S_j$ such that for all $i,j$ the crosscorrelations $C_{S_i, S_j}(d, \alpha, \beta)$ are small relative to the length of the sequences. The following theorem gives such a family.

Proposition 6.3.2 Let $E$ be an elliptic curve defined over the finite field $\mathbb{F}_q$ of characteristic $p$, given by a Weierstrass equation. Let $D$ be a positive integer, relatively prime to $p$, such that $k(E) | D | \#E(\mathbb{F}_q)/k(E)$. Let $P$ be a generator of the group $[D]E(\mathbb{F}_q)$. Let $f_1$ and $f_2$ be two polynomials in the coordinate functions $x \text{ and } y$ such that deg$_y(f_i) \leq 1$ for $i = 1, 2$ and such that for all $(\alpha, \beta) \in \mathbb{F}_q^2 - \{0, 0\}$ we have $p \nmid \text{deg}_x(\alpha f_1 - \beta f_2)$. Write $S_1 = S_{f_1, P}$ and $S_2 = S_{f_2, P}$. For all $\alpha, \beta \in \mathbb{F}_q^*$ and $0 \leq d < \#P > 0$ we have

$$|C_{S_1, S_2}(d, \alpha, \beta)| \leq \frac{1}{\#P} \left( 2 + \frac{D}{k(E)} \left( 2 + \text{deg}(f) + \text{deg}(g) \right) \sqrt{q^r} \right),$$

unless $d = 0$ and $\alpha f_1 = \beta f_2$.

Proof: This is a straightforward generalization of the proof of Theorem 6.2.19.

\[ \square \]

Theorem 6.3.3 Let $E$ be an elliptic curve defined over the finite field $\mathbb{F}_2$. Let $D$ be an odd positive integer such that $k(E) | D | \#E(\mathbb{F}_2)/k(E)$ and denote by $P$ a generator of the group $[D]E(\mathbb{F}_2)$. Let $S_a$ be the basic binary sequence $S_{f_{a, P}}$ with defining function $f = a_0 y + \cdots + a_n x^n y$, where $a_i \in \mathbb{F}_2$. For any number $0 \leq d < \#P > 0$ and $a, b \in \mathbb{F}_2^{n+1} - \{0\}$ we have

$$|C_{S_a, S_b}(d)| \leq \frac{1}{\#P} \left( 2 + \frac{D}{k(E)} \left( 2 + \text{deg}(f_a) + \text{deg}(f_b) \right) \sqrt{2^r} \right) \leq \frac{1}{\#P} \left( 2 + 4 \frac{D}{k(E)} (n + 2) \right),$$

unless $d = 0$ and $a = b$.

Proof: This is clear from the previous proposition. \[ \square \]
References


125
Fermat, class field theory and complex multiplication.


Index

\textit{p}-degenerate, 28

arithmetic length, 20
Artin-Schreier extension, 6, 7
auto-correlation, 115

Baker’s theorem, 11, 12, 24, 57
balance, 114
basic differential form, 55, 66
basic sequence, 115
blow-up, 17

cross-correlation, 124
curve of type I, 61
curve of type II, 61
delta-invariant, 11, 12, 16, 28
design, 12
desingularization tree, 18, 27

Euler’s equation, 22
exceptional curve, 17

first Garcia-Stichtenoth tower, 65, 68
Frobenius eigenvalues, 4
Frobenius map, 111
gap, 49
genus, 2, 12, 24, 36, 49

Hasse-Weil bound, 4
Hermitian curve, 38, 39

Ihara bound, 5
infinitely near point, 18
integrally indecomposable, 13
intersection

multiplicity, 16
isolated edge, 28

Klein quartic, 11, 38, 42, 47, 83
Kummer’s theorem, 15

leaves, 18

main part, 19
Milnor number, 24
Milnor’s equation, 24
Minkowski-sum, 13
missing functions, 83

Newton

diagram, 12, 19
polygon, 11–13
polytope, 12, 13
non-degenerate

w.r.t. Newton diagram, 22, 23
w.r.t. Newton polygon, 24, 54,
63, 64

Oesterlé bound, 5
one-point AG codes, 2

Plücker’s formula, 12, 17
plane curve, 53
prime-generating elliptic curve, 91

ramification divisor, 6
ramification index, 51
resolution tree, 18

second Garcia-Stichtenoth tower, 79
semigroup, 47, 58
special point, 18
strict transform, 18
substitute-and-reduce construction, 11, 13
Suzuki curve, 12, 32, 43

trace, 4
trinomial, 32
trinomial curve, 32

Weierstrass semigroup, 47
weighted degree, 14
Samenvatting

In dit proefschrift wordt onderzoek gedaan naar een aantal onderwerpen dat tussen algebraïsche meetkunde en coderingstheorie en cryptografie inzit. In het eerste hoofdstuk wordt een inleiding gegeven in de connectie tussen algebraïsche meetkunde en coderingstheorie zoals die beschreven is door Goppa. Verder geven de aandacht aan een aantal algebraïsche stellingen zoals die van Riemann, Hasse-Weil en Hurwitz. Het eerste hoofdstuk besluit met het geven van de expliciete beschrijving van twee asymptotische goede torens zoals gedaan door Garcia en Stichtenoth in (Garcia and Stichtenoth 1995; Garcia and Stichtenoth 1996) en het schetsen van een aantal onderzoeksfragen die nog open zijn op dit gebied.

Het tweede hoofdstuk behandelt de theorie van Newtonpolygone voor vlakke algebraïsche krommen. Deze theorie is bekend voor krommen gedefinieerd over lichamen met karakteristiek nul, maar wordt hier geëxperimenteerd naar karakteristiek ongelijk nul. Verder worden er enige toepassingen gegeven van deze theorie; met name worden trinomiale krommen geklassificeerd. Verder worden zogenaamde $p$-ontaaide singulariteiten onderzocht.

In het derde hoofdstuk wordt een verband tussen reguliere differentiaalvormen en gaten in een Weierstrasshalfgroup gegeven en uitgewerkt. In het geval van de zogenaamde krommen van type II levert dit, onder zekere technische restricties, een volledige beschrijving van een Weierstrasshalfgroup op. De theorie wordt ook toegepast op de zogenaamde eerste Garcia-Stichtenoth toren, in welk geval een gedeeltelijke beschrijving van de gewenste Weierstrasshalfgroup gevonden wordt.

Het vierde hoofdstuk keert weer terug naar krommen van type II, maar nu wordt, onder zekere technische restricties, een volledige beschrijving gegeven van de ring $K_{\infty}([0 : 1 : 0])$. Dit betekent dat voor deze krommen nu een expliciet recept bestaat om eenpunts Goppacodes met steunpunt $[0 : 1 : 0]$ te geven.

In het vijfde hoofdstuk wordt een meer cryptografische of getaltheoretische toepassing van elliptische krommen gegeven. Het blijkt dat met elliptische krommen families van Mersenne-achtige priemgetallen beschreven kunnen worden. Heuristieken die voor Mersennepriemgetallen bestaan, worden geëxperimenteerd naar deze families. Voor twee families geven we een expliciete simpele priemtest. Enkele grote priemgetallen zijn op deze manier gevonden.

Het zesde en laatste hoofdstuk geeft een andere toepassing van elliptische krommen. Een aantal manieren wordt beschreven om vanuit een elliptische kromme
zogenaamde pseudorandom rijen te vinden. De gegeven constructies worden onder-
zocht en resultaten over de correlatie-eigenschappen van de gevonden rijen, worden
gegeven.
Curriculum Vitae


Van 1996 tot 2001 was Peter werkzaam bij de Technische Universiteit Eindhoven als AIO, waar hij gewerkt heeft aan een NWO-project omtrent algebraïsch meetkundige codes.