Algebraic Curves and Codes

Ivan Soprunov

Department of Mathematics, Cleveland State University, 2121 Euclid Ave, Cleveland OH, USA
E-mail address: i.soprunov@csuohio.edu
Abstract.
## Contents

Preface vii  

Chapter 1. Introduction to Coding Theory 1  
1.1. Basic Definitions and First Examples 1  
1.2. Dual Codes 7  
1.3. Reed–Solomon Codes: Two constructions 9  
1.4. Cyclic Codes 12  
1.5. Asymptotic of Codes 14  
Exercises 19  

Chapter 2. Algebraic Curves 23  
2.1. Fields and Polynomial Rings 23  
2.2. Affine and Projective Curves 33  
2.3. Tangent Lines and Singular Points 40  
2.4. Bézout’s Theorem and Applications 45  
2.5. The Genus of a Curve 55  
Exercises 63  

Chapter 3. The Riemann–Roch Theorem 67  
3.1. Functions and Local Rings 67  
3.2. Divisors 74  
3.3. Differential Forms 79  
3.4. The Riemann–Roch Formula 83  
3.5. Elliptic Curves 89  
Exercises 95  

Chapter 4. Curves over Finite Fields 97  
4.1. Curves over non-algebraically closed fields 97  
4.2. The Zeta Function 99  
4.3. The Hasse–Weil Bound 99  
4.4. Hermitian Curves 99  

Chapter 5. Algebraic Geometry Codes 101  
5.1. The $L$-construction 101  
5.2. The $\Omega$-construction 101  
5.3. Duality 101  
5.4. (Quasi-)Self-Dual AG codes 101  
5.5. Asymptotics of AG codes 101
Preface
CHAPTER 1

Introduction to Coding Theory

The theory of error-correcting codes is a rapidly growing branch of Information theory. It employs various methods in modern mathematics ranging from probability and analysis to combinatorics and algebra. We will be studying methods of algebraic geometry (algebraic curves) in the theory of error-correcting codes. This is motivated by a recent discovery of algebraic geometry codes whose asymptotic behavior turned out to be better than of any previously known codes. What this really means we are going to find out in Section 1.5 and Section 5.5.

1.1. Basic Definitions and First Examples

The main problem in information theory is that data gets corrupted when transmitted. Digital data is a (very long) sequence of 0's and 1's, called bits. When a bit gets flipped (from 0 to 1 or vice versa) we say there is an error in this bit. Our goal is to (a) detect and (b) correct possible errors in data transmission. The basic idea in achieving this goal is adding redundancy to the data.

The following is a basic scheme of error correction:

message $\rightarrow$ encoded message $\rightarrow$ corrupted message $\rightarrow$ message

In the first step we add redundancy, in the second step the transmitted message gets errors, in the third step we detect and correct errors and remove redundancy to (hopefully) recover the original message. Here is an example of an error in the 3rd place:

$0110 \rightarrow 0100$

Here is a very simple way to add redundancy.

**Example 1.1.** Repetition code.

(a) Let us repeat every digit in the message twice:

$0110 \rightarrow 00111100 \rightarrow 00110100$.

Note that after the second step the digits in the third block are not the same. Hence we can say that there was an error either in the 5th and the 6th bit. However, we cannot say for sure whether it was 00 or 11 we transmitted. Therefore, this code can detect errors, but cannot correct them.

(b) Now we repeat every digit in the message three times:

$0110 \rightarrow 0001111000 \rightarrow 0011111000$.

Now we can see that the error occurred in the first block of three. Moreover, most likely the 001 came from 000, rather than 111. Therefore, we can say that this code can detect and correct up to one error per three bits.
1. Introduction to Coding Theory

In the previous examples we were applying the following maps:

\[(a) \quad 0 \rightarrow 00, \quad 1 \rightarrow 11\]
\[(b) \quad 0 \rightarrow 000, \quad 1 \rightarrow 111\]

We will call \(\{00, 11\}\) and \(\{000, 111\}\) the sets of codewords for the code in (a) and in (b), respectively.

**Question.** How many times do we need to repeat each digit to be able to correct two errors per codeword?

**Answer.** Five times. The codewords are \(\{00000, 11111\}\).

More generally, we need to repeat \(2^m + 1\) times to correct \(m\) errors per codeword.

### 1.1.1. The 4-7 Hamming Code.

The disadvantage of using a repetition code is that it is not very efficient. For example, the encoded message in the triple repetition code takes up 3 times as much space as the original message. We then will say that the efficiency of this code is \(1/3\). The following code will be able to correct up to 1 error per codeword, yet will have efficiency \(4/7\!\!\!\!\!\.\)

**Example 1.2. 4-7 Hamming Code.** The idea is to encode each block of 4 digits into a block of 7 digits in the following way:

\[a_1a_2a_3a_4 \mapsto a_1a_2a_3a_4a_5a_6a_7,\]

where

\[a_5 = a_1 + a_2 + a_3 \mod 2, \quad a_6 = a_1 + a_3 + a_4 \mod 2, \quad a_7 = a_2 + a_3 + a_4 \mod 2.\]

For example, \(0110 \mapsto 0110010\).

**Question.** How many codewords does this code have?

**Answer.** Same as the number of blocks of length 4, i.e. \(2^4\).

The correction can go as follows: for every received word look at the closest codeword, i.e. the one that differs from the word in at most one place (we’ll say “one flip away”).

**Proposition 1.3.** The 4-7 Hamming code has the following properties.

1. It can correct up to 1 error per codeword,
2. It has efficiency \(4/7\),
3. Every received word can be corrected.

**Proof.** We will prove (1) later. The idea is that you need at least 3 flips to get from one codeword to another. Part (2) is true since we need 7 bits for every 4 bits of the message. For part (3) note that for every codeword there are 7 words that are one flip away from it. Also no word can be one flip away from two distinct codewords, by part (1). Hence the total number of words that can be corrected is \(2^4 + 7 \cdot 2^4 = 2^7\), which is the total number of all possible words of length 7. \(\square\)
1.1. Using Linear Algebra. We will now use linear algebra over the field of two elements \( \mathbb{F}_2 = \{0, 1\} \). We will look at finite fields more closely in Section 2.1.1. Right now all we need to remember is that all operations (addition, multiplication, division) are modulo 2. Let us rewrite the definition of the 4-7 Hamming code in matrix notation. Its codewords \((a_1, \ldots, a_7) \in \mathbb{F}_7^2\) are computed by the following formula:

\[
(a_1, a_2, a_3, a_4, a_5, a_6, a_7) = (a_1, a_2, a_3, a_4)
\begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 1
\end{bmatrix}.
\]

Note that we use left multiplication, so our messages are composed of rows of length 4 and our codewords are rows of length 7. Let \( G \) denote the above matrix. Then it defines the following sequence of linear maps:

\[
0 \longrightarrow \mathbb{F}_2^4 \xrightarrow{G} \mathbb{F}_2^7 \xrightarrow{H} \mathbb{F}_2^3 \longrightarrow 0.
\]

This diagram is called a short exact sequence. What this simply means is that the map given by \( G \) is injective, the map given by \( H \) is surjective, and \( \text{Im}(G) = \text{Ker}(H) \). Such \( H \) is not unique (we will see later how to find such a matrix), and is called a parity check matrix. You should check that the following matrix is a valid choice of \( H \):

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

We will use \( H \) for detecting and correcting errors. By construction, the codewords are the elements of \( \text{Im}(G) = \text{Ker}(H) \). Hence \( c \) is a codeword iff \( cH = 0 \). Now suppose \( x \) is the word obtained from \( c \) by flipping its \( i \)th digit. Then \( x = c + e_i \), where \( e_i \) is the \( i \)th standard basis vector (remember that this is vector addition modulo 2). Hence \( xH = (c + e_i)H = e_iH \), which is the \( i \)th row of \( H \). We obtain the following decoding algorithm:

- **Input:** \( x \in \mathbb{F}_2^7 \)
- If \( xH = 0 \) then **Output:** \( c = x \)
- If \( xH = \text{i} \)th row of \( H \) then **Output:** \( c = x - e_i \).

**Example 1.4.**

(a) Let \( x = (1101000) \). Then

\[
xH = (1, 1, 0, 1, 0, 0, 0) \begin{bmatrix}
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & 1 & 1 \\
0 & 1 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} = (0, 0, 0).
\]

Thus \( x \) is a codeword, and no error occurred. The original message was (1101).
(b) Let \( x = (1111110) \). Then
\[
\begin{bmatrix}
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
1, 1, 1, 1, 1, 0
\end{bmatrix}
= (0, 0, 1),
\]
which is the 7th row of \( H \). Hence there was a flip in the 7th place, and the transmitted codeword was \((1111111)\). The original message was \((1111)\).

(c) Let \( x = (1000011) \). Then
\[
\begin{bmatrix}
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
1, 1, 1, 1, 1, 0
\end{bmatrix}
= (1, 0, 1),
\]
which is the 2nd row of \( H \). Hence there was a flip in the 2nd place, and the transmitted codeword was \((1100011)\). The original message was \((1100)\).

1.1.3. General Definitions. In general, we don’t want to restrict ourselves to sequences of 0’s and 1’s, we can consider words involving other “letters”. So let us fix a finite set \( A \), called an alphabet. By a word of length \( n \) we will mean any sequence of elements of \( A \) of length \( n \), i.e. an element of \( A^n = A \times \cdots \times A \). In all our later examples, though, \( A \) will be a finite field, such as integers mod a prime number. (More about finite fields will be in Section 2.1.1).

Definition 1.5. A code \( C \) is a subset of \( A^n \). Its elements \( c = (c_1, \ldots, c_n) \) are called codewords.

We can define the notion of the distance on the set \( A^n \). This is a convenient way of measuring how different two words are.

Definition 1.6. The Hamming distance between \( a \) and \( b \) in \( A^n \) is defined by
\[
\text{dist}(a, b) = \# \text{ of places } a \text{ differs from } b = \# \{ i \mid a_i \neq b_i \}.
\]

Now we will define the main parameters of a code. We set \( q = |A| \), the size of the alphabet.

Definition 1.7. Let \( C \subseteq A^n \) be a code. Define
- \( n \) to be the length of the code \( C \);
- \( k = \log_q(|C|) \) to be the dimension of the code \( C \);
- \( d = \min\{\text{dist}(a, b) \mid a, b \in C, a \neq b\} \) to be the minimum distance of \( C \).

A code over an alphabet \( A \) of size \( q \) with parameters \( n, k, d \) will be referred to as an \([n, k, d]_q\)-code.

Example 1.8. (a) The triple code: \( C = \{(0,0,0), (1,1,1)\} \subseteq \mathbb{F}_3^3 \). The parameters are \( n = 3, k = \log_2(2) = 1, \) and \( d = 3 \).
1.1. BASIC DEFINITIONS AND FIRST EXAMPLES

(b) The 4-7 Hamming code:
\[ C = \{(a_1, \ldots, a_7) \in \mathbb{F}_2^7 \mid a_5 = a_1 + a_2 + a_3, a_6 = a_1 + a_3 + a_4, a_7 = a_2 + a_3 + a_4 \}. \]

The parameters are \( n = 7, \ k = \log_2(2^4) = 4, \) and \( d = 3 \) (as we will see later).

When studying families of codes with increasing length the following relative parameters are useful.

**Definition 1.9.** Let \( C \subset A^n \) be a code. Define

- \( R = k/n \) to be the **information rate** of the code \( C \);
- \( \delta = d/n \) to be the **relative minimum distance** of \( C \).

Note that the information rate measures the efficiency of the code and the relative minimum distance measures the reliability of the code.

Almost all our future results will concern a special class of codes, called linear codes. In this case we need to set \( A = \mathbb{F}_q \), a finite field. Here is the definition.

**Definition 1.10.** A code \( C \subset \mathbb{F}_q^n \) is called **linear** if \( C \) is a vector space over \( \mathbb{F}_q \), i.e.

- (1) \( 0 \in C \);
- (2) \( C \) is closed under vector addition over \( \mathbb{F}_q \);
- (3) \( C \) is closed under scalar multiplication by elements of \( \mathbb{F}_q \).

**Example 1.11.**

(a) The triple code:
\[ C = \{(0,0,0),(1,1,1)\} = \text{span}_{\mathbb{F}_2}\{(1,1,1)\}. \]

Geometrically, this is a line in the direction of \((1,1,1)\) in the 3-dimensional space over \( \mathbb{F}_2 \).

(b) The 4-7 Hamming code:
\[ C = \{(a_1, \ldots, a_7) \in \mathbb{F}_2^7 \mid a_1 + a_2 + a_3 + a_5 = a_1 + a_3 + a_4 + a_6 = a_2 + a_3 + a_4 + a_7 = 0 \}. \]

This is the solution set of a linear system, hence a vector space. Geometrically, this is a 4-dimensional subspace, which is the intersection of three hyperplanes in \( \mathbb{F}_2^7 \).

There is no risk of confusing the dimension of a code and its dimension as a vector space, as they turn out to be equal.

**Proposition 1.12.** If \( C \) is a linear code then its dimension as a code equals its dimension as a vector space over \( \mathbb{F}_q \).

**Proof.** Let \( k = \dim_{\mathbb{F}_q} C \). Then we can choose a basis \( B = \{v_1, \ldots, v_k\} \) for \( C \). Every \( c \) has a unique representation as a linear combination \( c = \lambda_1 v_1 + \cdots + \lambda_k v_k \), for \( \lambda_i \in \mathbb{F}_q \). Thus the number of elements in \( C \) equals \( q^k \) (there are \( q \) choices for every coefficient \( \lambda_i \)). Therefore, \( \log_q(|C|) = \log_q(q^k) = k \), as stated. \( \square \)

Notice that the rows of \( G \) from Section 1.1.2 form a basis for the 4-7 Hamming code.

**Definition 1.13.** A matrix \( G \) whose rows form a basis for a linear code \( C \) is called a **generator matrix** of \( C \).

For every codeword in \( C \) we can measure its distance to 0. This is called the weight of the codeword. Here is an equivalent definition.
Definition 1.14. Let C be a linear code and \( c \in C \) a codeword. Define the weight of \( c \) as
\[
 w(c) = \# \text{ of non-zero entries in } c.
\]
The minimum weight of \( C \) is the smallest weight of all non-zero codewords in \( C \).

Proposition 1.15. Let \( C \) be a linear code. Then its minimum distance equals its minimum weight.

Proof. Indeed, we have
\[
d = \min \{ \text{dist}(a, b) \mid a, b \in C, a \neq b \} = \min \{ \text{dist}(c, 0) \mid c \in C, c \neq 0 \} = \min \{ w(c) \mid c \in C, c \neq 0 \}.
\]

Here are a few simple examples of linear codes.

Example 1.16. “Trivial” codes:
(a) Let \( C = \mathbb{F}_q^n \). Clearly, \( k = n \) and \( d = 1 \) as \((1, 0, \ldots, 0)\) is a vector of smallest possible weight. Thus, this is a \([n, n, 1]_q\)-code. For a generator matrix we can take \( G = I_n \), the \( n \times n \) identity matrix.
(b) Let \( C = \{ (a, \ldots, a) \in \mathbb{F}_q^n \mid a \in \mathbb{F}_q \} \). This is a \([n, 1, n]_q\)-code with a generator matrix \( G = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \).
(c) Let \( C = \{ (a_1, \ldots, a_n) \in \mathbb{F}_q^n \mid \sum_{i=1}^n a_i = 0 \} \). Note that no non-zero codeword can have weight 1, and \((1, 0, \ldots, 0, -1)\) is a codeword of weight 2. Therefore, \( d = 2 \). We get an \([n, n-1, 2]_q\)-code. For a generator matrix we can take
\[
 G = \begin{bmatrix} 1 & 0 & \cdots & -1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & -1 \end{bmatrix}
\]

1.1.4. Spheres and Balls. Let us return to the general situation. Let \( A \) be an alphabet and \( A^n \) the set of words of length \( n \). As you can check (see Exercise 1.2) the Hamming distance defines a metric on \( A^n \). What do spheres and balls in this metric space look like?

Definition 1.17. A sphere in \( A^n \) with center \( c \) and radius \( r \) is the set
\[
 S(c, r) = \{ a \in A^n \mid \text{dist}(a, c) = r \}.
\]
Similarly, a ball in \( A^n \) with center \( c \) and radius \( r \) is the set
\[
 B(c, r) = \{ a \in A^n \mid \text{dist}(a, c) \leq r \}.
\]

Proposition 1.18. The number of elements in a sphere and in a ball are given by
\[
|S(c, r)| = \binom{n}{r}(q-1)^r;
\]
\[
|B(c, r)| = \sum_{i=0}^r \binom{n}{i}(q-1)^i.
\]
1.2. Dual Codes

Proof. (1) Since dist(a, c) = r there are exactly r entries where a differs from c. To count all such a, first note that there are \( \binom{n}{r} \) ways to choose r entries where a would differ from c. Second, at each of these r entries we can place any of the q – 1 values for a
t that are not equal to c_1.

(2) This follows from the observation that \( B(c, r) \) is the disjoint union of spheres of radii \( i = 0, 1, \ldots, r \).

The following theorem relates the minimum distance of a code to the number of errors per codeword it can correct.

**Theorem 1.19.** Let \( C \) be a code with minimum distance \( d \). Then \( C \) can correct up to \( \left\lfloor \frac{d-1}{2} \right\rfloor \) errors per codeword.

Proof. Consider the ball \( B(c, r) \) of radius r around each codeword \( c \in C \). If \( r = \left\lfloor \frac{d-1}{2} \right\rfloor \) then the balls are disjoint. If there were less than or equal to \( \left\lfloor \frac{d-1}{2} \right\rfloor \) errors in a codeword then the resulting word a belongs to exactly one of the balls. Hence we can recover the codeword by looking at the center of the ball containing a.

### 1.2. Dual Codes

Recall that for any subspace of a vectors space (equipped with a dot product) we can consider its orthogonal complement. This brings us to the notion of the dual to a linear code.

Let \( a, b \) be two vectors in \( \mathbb{F}_q^n \). Their dot product is defined by \( a \cdot b = \sum_{i=1}^{n} a_i b_i \). Note that this is an element of \( \mathbb{F}_q \). When \( a \cdot b = 0 \) we say that \( a \) and \( b \) are orthogonal.

**Definition 1.20.** Let \( C \subseteq \mathbb{F}_q^n \) be a linear code. The dual code \( C^\perp \) is defined by

\[ C^\perp = \{ a \in \mathbb{F}_q^n \mid a \cdot b = 0 \text{ for any } b \in C \}. \]

The next properties of dual codes follow from standard facts in Linear Algebra.

**Proposition 1.21.** Let \( C \subseteq \mathbb{F}_q^n \) be a linear code. Then

1. \( (C^\perp)^\perp = C \);
2. \( \dim(C^\perp) = n - \dim(C) \).

**Example 1.22.** The “trivial” codes (b) and (c) from Example 1.16 are dual to each other. Let \( C_1 \) denote the code in (b) and \( C_2 \) denotes the code in (c). It is easy to check that every codeword in \( C_1 \) is orthogonal to every codeword in \( C_2 \). Hence \( C_1 \subseteq C_2^\perp \). On the other hand, \( \dim(C_1) = 1 \) and \( \dim(C_2^\perp) = n - (n - 1) = 1 \) by part (2) of Proposition 1.21. Therefore, \( C_1 = C_2^\perp \).

The situation over a finite field is a little different than what you are used to when looking at orthogonal vectors in \( \mathbb{R}^n \). For example, there exist non-zero self-orthogonal vectors in \( \mathbb{F}_q^n \).

**Example 1.23.** Let \( \mathbb{F}_3 = \{0, 1, 2\} \) be the field of integers mod 3. The vector \( (1, 1, 1) \in \mathbb{F}_3^n \) is orthogonal to itself. Consider

\[ C = \text{span}_{\mathbb{F}_3}\{(1, 1, 1)\} = \{(0, 0, 0), (1, 1, 1), (2, 2, 2)\}. \]

Then \( C \subseteq C^\perp \). What is \( C^\perp \)? By Proposition 1.21, \( C^\perp \) is 2-dimensional. Also \( (1, 2, 0) \in C^\perp \). Therefore,

\[ C^\perp = \text{span}_{\mathbb{F}_3}\{(1, 1, 1), (1, 2, 0)\}. \]
Note that $C^\perp$ consists of 9 vectors. Can you list them all?

Next we will see how to find a generator matrix for $C^\perp$, given a generator matrix for $C$. Let $C$ be a linear $[n, k, d]_q$-code with generator matrix $G$. Consider a short exact sequence

$$0 \rightarrow \mathbb{F}_q^k \xrightarrow{G} \mathbb{F}_q^n \xrightarrow{H} \mathbb{F}_q^{n-k} \rightarrow 0,$$

where $H$ is a parity check matrix for $C$ (see Section 1.1.2).

We have the following proposition.

**Proposition 1.24.** Let $H$ be a parity check matrix for $C$. Then $H^t$ is a generator matrix for $C^\perp$.

**Proof.** By taking the transpose of $G$ and $H$ we obtain a dual sequence (note that the maps are now reversed).

$$0 \rightarrow \mathbb{F}_q^{n-k} \xrightarrow{H^t} \mathbb{F}_q^n \xrightarrow{G^t} \mathbb{F}_q^k \rightarrow 0,$$

which is also exact (check that). Hence

$${\text{Im}}(H^t) = {\text{Ker}}(G^t) = \{a \in \mathbb{F}_q^n | aG^t = 0\} = \{a \in \mathbb{F}_q^n | a \cdot b = 0 \text{ for each column } b \text{ of } G^t\}.$$

But the columns of $G^t$ are the rows of $G$ and they span $C$. Hence, the latter equals

$$\{a \in \mathbb{F}_q^n | a \cdot b = 0 \text{ for every } b \in C\} = C^\perp.$$

Therefore, $\text{Im}(H^t) = C^\perp$ and so the rows of $H^t$ span $C^\perp$. Also the rows of $H^t$ are linearly independent since the map $H^t$ is injective. In other words, the rows of $H^t$ is a basis for $C^\perp$. □

This proof also tells us how one can compute $H$. Consider the homogeneous linear system $Gx = 0$. From linear algebra we know how to write down a basis $\{b_1, \ldots, b_{n-k}\}$ for its solution space (i.e. a basis for the null space of $G$). They are the columns of $H = [b_1, \ldots, b_{n-k}]$.

**Example 1.25.** Consider a linear code $C$ with a generator matrix $G = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$.

The null space of $G$ is spanned by $(1, 1, 1)$, so $H = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is a parity check matrix.

By Proposition 1.24, $H^t = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ is a generator matrix for the dual code $C^\perp$. Note that this is a particular case of Example 1.22. The following are the two short exact sequences for $C$ and $C^\perp$:

$$0 \rightarrow \mathbb{F}_q^2 \xrightarrow{\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}} \mathbb{F}_q^3 \xrightarrow{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}} \mathbb{F}_q^1 \rightarrow 0,$$

$$0 \rightarrow \mathbb{F}_q^1 \xrightarrow{\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}} \mathbb{F}_q^3 \xrightarrow{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 \end{bmatrix}} \mathbb{F}_q^2 \rightarrow 0,$$
1.3. Reed–Solomon Codes: Two constructions

We now turn to a very important example of a linear code constructed using spaces of polynomial functions. This is our first step towards studying algebraic geometry codes.

As usual, \( F[x] \) will denote the ring of polynomials over a field \( F \) in one variable \( x \). For the rest of the section we fix a subset \( \mathcal{P} = \{ \alpha_1, \ldots, \alpha_n \} \) of \( n \geq 1 \) elements of \( \mathbb{F}_q \), and fix an integer \( m < n \).

1.3.1. The \( \mathcal{L} \)-construction. Define

\[
\mathcal{L}(m) = \{ f \in \mathbb{F}_q[x] \mid \deg f \leq m \}.
\]

Note that \( \mathcal{L}(m) \) is a vector space over \( \mathbb{F}_q \) of dimension \( m + 1 \). It has a basis of monomials \( \{ 1, x, x^2, \ldots, x^m \} \).

Next we define the evaluation map

\[
\text{ev}_\mathcal{P} : \mathcal{L}(m) \rightarrow \mathbb{F}_q^n, \quad f \mapsto (f(\alpha_1), \ldots, f(\alpha_n)).
\]

It satisfies the following properties.

**Proposition 1.26.** The evaluation map \( \text{ev}_\mathcal{P} \) is a linear map which is injective if \( m < n \).

**Proof.** We leave it to you to check that \( \text{ev}_\mathcal{P} \) is a linear map. Let us show that if \( m < n \) then \( \text{Ker}(\text{ev}_\mathcal{P}) = \{0\} \). We have

\[
\text{Ker}(\text{ev}_\mathcal{P}) = \{ f \in \mathcal{L}(m) \mid f(\alpha_1) = \cdots = f(\alpha_n) = 0 \},
\]

hence, if \( f \in \text{Ker}(\text{ev}_\mathcal{P}) \) then \( f \) has \( n \) distinct root. But any non-zero polynomial in \( \mathcal{L}(m) \) has at most \( m < n \) roots. Thus \( \text{Ker}(\text{ev}_\mathcal{P}) = \{0\} \), and so \( \text{ev}_\mathcal{P} \) is injective. \( \square \)

**Definition 1.27.** The image of the evaluation map \( \text{Im}(\text{ev}_\mathcal{P}) \) is called the Reed–Solomon code. We denote it by \( C_{m,\mathcal{P}} \).

In the next theorem we compute the parameters of the Reed–Solomon code.

**Theorem 1.28.** The Reed–Solomon code \( C_{m,\mathcal{P}} \) is an \([ n, m + 1, n - m ]_q\)-code.

**Proof.** The length is \( n \) by definition. By Proposition 1.26,

\[
\dim C_{m,\mathcal{P}} = \dim \text{Im}(\text{ev}_\mathcal{P}) = \dim \mathcal{L}(m) = m + 1.
\]

To show that \( d = n - m \) note again that every \( f \in \mathcal{L}(m) \) has at most \( m \) roots, so every codeword \( (f(\alpha_1), \ldots, f(\alpha_n)) \) has at least \( n - m \) non-zero entries. Also \( f(x) = (x - \alpha_1) \cdots (x - \alpha_m) \) lies in \( \mathcal{L}(m) \) and produces a codeword of weight exactly \( n - m \). Thus \( n - m \) is the minimum weight (minimum distance) of \( C_{m,\mathcal{P}} \). \( \square \)

Let us now write down a generator matrix for \( C_{m,\mathcal{P}} \). Since \( \text{ev}_\mathcal{P} \) is injective, the image of the monomial basis \( \{1, x, x^2, \ldots, x^m\} \) under \( \text{ev}_\mathcal{P} \) forms a basis for \( \text{Im}(\text{ev}_\mathcal{P}) \), i.e. \( \{\text{ev}_\mathcal{P}(1), \text{ev}_\mathcal{P}(x), \ldots, \text{ev}_\mathcal{P}(x^m)\} \) are the rows of a generator matrix \( G \). We obtain

\[
G = \begin{bmatrix}
1 & 1 & \ldots & 1 \\
\alpha_1 & \alpha_2 & \ldots & \alpha_n \\
\alpha_1^2 & \alpha_2^2 & \ldots & \alpha_n^2 \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_1^m & \alpha_2^m & \ldots & \alpha_n^m
\end{bmatrix}.
\]
Remark 1.29. When \( m = n - 1 \) we obtain a square Vandermonde matrix, familiar to you from linear algebra. Its determinant equals \( \prod_{1 \leq i < j \leq n} (\alpha_i - \alpha_j) \), which is non-zero since the \( \alpha_i \) are disjoint.

Example 1.30. Consider a Reed–Solomon code over \( \mathbb{F}_5 = \{0, 1, 2, 3, 4\} \) (integers mod 5) with \( n = 5 \), \( m = 2 \), and \( \mathcal{P} = \mathbb{F}_5 \). The space \( \mathcal{L}(2) \) has a basis \( \{1, x, x^2\} \). Therefore, the code \( C_{2, \mathbb{F}_5} \) has a generator matrix

\[
G = \begin{bmatrix}
1 & 1 & 1 & 1 \\
0 & 1 & 2 & 3 \\
0 & 1 & 4 & 4
\end{bmatrix}.
\]

By Theorem 1.28 this is a \([5, 3, 3]_5\)-code.

1.3.2. The \( \Omega \)-construction. As before, let \( \mathcal{P} = \{\alpha_1, \ldots, \alpha_n\} \) be a subset of \( \mathbb{F}_q \). Denote \( f_0(x) = (x - \alpha_1) \cdots (x - \alpha_n) \). Consider the set

\[
\Omega(m) = \left\{ \frac{g(x)}{f_0(x)} \mid g \in \mathbb{F}_q[x], \deg g \leq n - m - 2 \right\}.
\]

Note that \( \Omega(m) \) is a vector space over \( \mathbb{F}_q \) whose elements are rational functions with poles in \( \mathcal{P} \). The dimension of \( \Omega(m) \) equals the dimension of polynomials of degree up to \( n - m - 2 \), so \( \dim \Omega(m) = n - m - 1 \).

Definition 1.31. Let \( h \in \mathbb{F}_q[x] \). Define the residue of \( \frac{h}{f_0} \) at \( \alpha_i \) by

\[
\text{res}_{\alpha_i} \left( \frac{h}{f_0} \right) = \frac{(x - \alpha_i)h(x)}{f_0(x)} \bigg|_{x=\alpha_i} = \frac{h(\alpha)}{\prod_{j=1, j \neq i}^{n} (\alpha_i - \alpha_j)}.
\]

Example 1.32. Let us compute the residue of \( \frac{x^2 + 1}{x^2 - 1} \) at \( x = 1 \):

\[
\text{res}_1 \left( \frac{x^2 + 1}{x^2 - 1} \right) = \left. \frac{x^2 + 1}{x + 1} \right|_{x=1} = 1.
\]

Theorem 1.33. (The Residue Formula) If \( \deg h \leq n - 2 \) then

\[
\sum_{\alpha_i \in \mathcal{P}} \text{res}_{\alpha_i} \left( \frac{h}{f_0} \right) = 0.
\]

Before we prove this formula, let us remark that over the complex numbers it follows from the Cauchy residue formula. Indeed, by the Cauchy residue formula the sum of the residues over the \( \alpha_i \) equals negative of the residue at infinity. However, when \( \deg h \leq n - 2 \) the form \( \frac{h(x)}{f_0(x)} \) \( dx \) has no pole at infinity, so the residue there is zero.

Proof. Let \( h(x) = a_0 + a_1 x + \cdots + a_{n-2} x^{n-2} \). The following determinant is zero:

\[
\begin{vmatrix}
1 & 1 & \cdots & 1 \\
\alpha_1 & \alpha_2 & \cdots & \alpha_n \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_1^{n-2} & \alpha_2^{n-2} & \cdots & \alpha_n^{n-2} \\
\text{res}_{\alpha_1} & \text{res}_{\alpha_2} & \cdots & \text{res}_{\alpha_n}
\end{vmatrix} = 0.
\]
Indeed, the last row of the above matrix is a linear combination of the first \( n-1 \) rows with coefficients \( a_0, \ldots, a_{n-2} \). On the other hand, expanding the determinant along the last row and using Remark 1.29 we obtain

\[
0 = \sum_{i=1}^{n} (-1)^{n+i} h(\alpha_i) \prod_{j,k \neq i, j < k} (\alpha_k - \alpha_j) = \sum_{i=1}^{n} (-1)^{n} h(\alpha_i) \prod_{j \neq i} (\alpha_k - \alpha_j).
\]

Since the product in the numerator is independent of \( i \), we can factor it out:

\[
0 = (-1)^{n} \prod_{j < k} (\alpha_k - \alpha_j) \sum_{i=1}^{n} h(\alpha_i) \prod_{j \neq i} (\alpha_k - \alpha_j).
\]

It remains to notice that the sum on the right is precisely the sum of the residues in the residue formula. \( \square \)

Let us know return to the \( \Omega \)-construction of a Reed–Solomon code. Define the residue map

\[
\text{res}_P : \Omega(m) \to \mathbb{F}_q^n, \quad \frac{g}{f_0} \mapsto \left( \text{res}_{\alpha_1} \left( \frac{g}{f_0} \right), \ldots, \text{res}_{\alpha_n} \left( \frac{g}{f_0} \right) \right).
\]

**Proposition 1.34.** The residue map \( \text{res}_P \) is a linear map which is injective if \( m \geq 0 \).

**Proof.** It is easy to check that \( \text{res}_P \) is a linear map, so it is left it you. As in the proof of Proposition 1.26 we just need to show that \( \text{Ker}(\text{res}_P) = \{0\} \). Indeed, if \( 0 \neq g/f_0 \in \text{Ker}(\text{res}_P) \) then \( \text{res}_{\alpha_i} (g/f_0) = 0 \), for every \( 1 \leq i \leq n \). This implies that \( g(\alpha_i) = 0 \), for every \( 1 \leq i \leq n \) (see Definition 1.31). But then \( g \) would have \( n \) distinct roots, which is impossible since \( \text{deg} g \leq n - m - 2 < n \). Therefore \( g = 0 \) and \( \text{Ker}(\text{res}_P) = \{0\} \). \( \square \)

**Definition 1.35.** The image of the residue map \( \text{res}_P \) is called a dual Reed–Solomon code and denoted by \( C_{m,P}^* \).

The following theorem justifies the above definition.

**Theorem 1.36.** The code \( C_{m,P}^* \) is a \([n, n-m-1, m+2]_q\)-code, dual to the code \( C_{m,P} \).

**Proof.** First, by Proposition 1.34, \( \dim C_{m,P}^* = \dim \Omega(m) = n-m-1 \). To compute the minimum distance, consider a codeword corresponding to \( g/f_0 \in \Omega(m) \). As we saw in the proof of Proposition 1.34, the number of zero entries in the codeword equals the number of roots of \( g \), which cannot exceed \( n - m - 2 \). Therefore, the weight of the codeword is no less than \( n - (n - m - 2) = m + 2 \). On the other hand, it is easy to check that the polynomial \( g(x) = (x - \alpha_1) \cdots (x - \alpha_{n-m-2}) \) produces a codeword with weight exactly \( m + 2 \).

For the second part of the statement, let \( f \in \mathcal{L}(m) \) and \( g/f_0 \in \Omega(m) \). We will show that they define orthogonal codewords. Indeed, the dot product of \( (f(\alpha_1), \ldots, f(\alpha_n)) \) and \( \left( \text{res}_{\alpha_1} \left( \frac{g}{f_0} \right), \ldots, \text{res}_{\alpha_n} \left( \frac{g}{f_0} \right) \right) \) equals

\[
\sum_{i=1}^{n} f(\alpha_i) \text{res}_{\alpha_i} \left( \frac{g}{f_0} \right) = \sum_{i=1}^{n} \text{res}_{\alpha_i} \left( \frac{f g}{f_0} \right).
\]

But the latter sum equals zero by the Residue Formula, since \( \text{deg}(f g) = \text{deg} f + \text{deg} g \leq m + n - m - 2 = n - 2 \). This shows that \( C_{m,P}^* \subseteq C_{m,P}^\perp \). To see that they
are in fact the same, compare the dimensions: \( \dim C_{m,p}^* = n - (m + 1) = n - m - 1 \) by Proposition 1.21 and Theorem 1.28. Also \( \dim C_{m,p} = n - m - 1 \) by above.

**Example 1.37.** Consider a dual Reed–Solomon code over \( \mathbb{F}_5 = \{0, 1, 2, 3, 4\} \) with \( n = 5, m = 2, \) and \( P = \mathbb{F}_5. \) The space \( \Omega(2) \) consists of rational functions \( g/f_0, \) where \( f_0(x) = x(x-1)(x-2)(x-3)(x-4) \) and \( \deg g \leq n - m - 2 = 1. \) Hence \( \Omega(2) \) has a basis \( \{1/f_0, x/f_0\}. \) We leave it as an exercise (see Exercise 1.3) to check that the residue of \( 1/f_0 \) at every point of \( \mathbb{F}_5 \) equals \( 4. \) This produces a generator matrix for \( C_{2,\mathbb{F}_5}^* \):

\[
G^* = \begin{bmatrix}
4 & 4 & 4 & 4 & 4 \\
0 & 4 & 3 & 2 & 1
\end{bmatrix}.
\]

By Theorem 1.28 this is a \([5, 2, 4]_5\)-code. You can check that the transpose of \( G^* \) is a parity check matrix for the code in Example 1.30, hence the two codes are dual of each other.

### 1.4. Cyclic Codes

In this section we will look at a class of linear codes called cyclic. They are closely related to ideals of certain quotient rings. It turns out that some of the Reed–Solomon codes are examples of cyclic codes, as we will prove at the end of the section. The definition is simple.

**Definition 1.38.** A linear code \( C \subseteq \mathbb{F}_q^n \) is called **cyclic** if \((c_0, \ldots, c_{n-2}, c_{n-1}) \in C\) implies that \((c_{n-1}, c_0, \ldots, c_{n-2}) \in C.\)

In other words, a cyclic code is a linear code which is closed under a cyclic permutation of the entries in its codewords.

**Example 1.39.** Let \( C \) be a binary code (i.e. a code over \( \mathbb{F}_2 \)) with generator matrix

\[
G = \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & 0 & 1 & 0
\end{bmatrix}.
\]

This is a cyclic code. Indeed, we can list the elements of \( C: \)

\[
C = \{(0, 0, 0, 0), (1, 1, 1, 1), (1, 0, 1, 0), (0, 1, 0, 1)\}.
\]

Now it is easy to see that this set is closed under a cyclic permutation.

Next we will see how cyclic codes are related to ideals in quotients of polynomial rings. We are going to associate to every codeword \( c = (c_0, c_1, \ldots, c_{n-1}) \) in \( C \) a polynomial \( c(t) = c_0 + c_1 t + \cdots + c_{n-1} t^{n-1} \) in \( \mathbb{F}_q[t]. \) This way the code \( C \) corresponds to a set of polynomials

\[
I_C = \{c(t) \in \mathbb{F}_q[t] \mid c \in C\}.
\]

Note that

\[
tc(t) = c_0 t + c_1 t^2 + \cdots + c_{n-1} t^n \equiv c_{n-1} + c_0 t + \cdots + c_{n-2} t^{n-1} \mod (t^n - 1).
\]

Hence, the polynomial corresponding to the cyclic permutation of \( c \) is obtained by reducing \( tc(t) \) modulo \( t^n - 1. \) This suggests that we need to consider the image of \( I_C \) in the quotient ring \( \mathbb{F}_q[t]/(t^n - 1). \) We will keep the same notation:

\[
I_C = \{c(t) \in \mathbb{F}_q[t]/(t^n - 1) \mid c \in C\}.
\]
EXAMPLE 1.40. Let us look at the quotient ring $R = \mathbb{F}_2[t]/(t^4 - 1)$. The remainders mod $t^4 - 1$ have degree up to three, so we can write

$$R = \{a_0 + a_1 t + a_2 t^2 + a_3 t^3 \mid a_i \in \mathbb{F}_2\} = \{0, 1, t, 1 + t, t^2, 1 + t^2, \ldots\},$$

where we should keep in mind that these are classes of polynomials mod $t^4 - 1$. For example, $(1 + t^2)t^2 = t^2 + t^4 = 1 + t^2$ in $R$.

Consider the ideal $I$ in $R$ generated by $1 + t^2$, i.e. $I = (1 + t^2) = \{h(t)(1 + t^2)\}$. Since when writing elements of $R$ we don’t need to consider polynomials of degree greater than 3, it is enough to take all $h(t)$ of degree up to 1, so we can write

$$I = \{h(t)(1+g^2)\} = \{0, 1+t^2, t(1+t^2), (1+t)(1+t^2)\} = \{0, 1+t^2, t+t^3, 1+t+t^2+t^3\}.$$

Note that the coefficients of these polynomials are the vectors

$$\{(0,0,0,0), (1,0,1,0), (0,1,0,1), (1,1,1,1)\},$$

which is precisely the cyclic code $C$ in Example 1.39. Hence $I = I_C$.

This correspondence between ideals in $\mathbb{F}_q[t]/(t^n - 1)$ and cyclic codes over $\mathbb{F}_q$ persists in general.

THEOREM 1.41. $C \subseteq \mathbb{F}_q^n$ is a cyclic code of length $n$ if and only if $I_C$ is an ideal in $\mathbb{F}_q[t]/(t^n - 1)$.

PROOF. ($\Leftarrow$) If $I_C$ is an ideal in $R = \mathbb{F}_q[t]/(t^n - 1)$ then it is closed under addition and multiplication by elements in $R$. In particular, $tI_C \subseteq I_C$. This means that whenever $c(t) \in I_C$ then also $tc(t) \in I_C$. But we saw that this is equivalent to $(c_{n-1}, c_0, \ldots, c_{n-2}) \in C$ whenever $(c_0, \ldots, c_{n-2}, c_{n-1}) \in C$, i.e. $C$ is cyclic.

($\Rightarrow$) Since $C$ is linear, the set $I_C$ forms a subgroup of $R$ under addition and is closed under multiplication by elements of $\mathbb{F}_q$. If $C$ is cyclic then it is closed under the cyclic permutation of the entries in its codewords. This implies that $I_C$ is closed under multiplication by $t \in R$, i.e. $tI_C \subseteq I_C$. But then $t^i I_C \subseteq I_C$ for any $i \geq 1$. This implies that $f I_C \subseteq I_C$ for any $f \in R$, i.e. $I_C$ is an ideal. 

It is a standard fact in abstract algebra that the ring of polynomials $F[t]$ over a field $F$ is a PID (principal ideal domain). This means that every ideal $I$ in $F[t]$ is generated by a single element, $I = \langle g \rangle$, for some $g \in F[t]$. The same is true about quotient rings $R = F[t]/J$ where $J$ is an ideal in $F[t]$. In fact, any ideal $I$ in $R$ corresponds to a unique ideal $I' \supseteq J$ in $F[t]$. Now if $I'$ is generated by some $g' \in F[t]$, then $I$ is generated by the image of $g'$ in $R$. Therefore $R$ is also a PID. In our situation, this implies that for any cyclic code $C$ there exists a polynomial $g_C$ whose image in $\mathbb{F}_q[t]/(t^n - 1)$ generates $I_C$.

DEFINITION 1.42. Let $C$ be a cyclic code. The monic polynomial $g_C$ whose image in $\mathbb{F}_q[t]/(t^n - 1)$ generates $I_C$ is called the generator polynomial for $C$.

EXAMPLE 1.43. The generator polynomial for the cyclic code in Example 1.39 is $1 + t^2$.

We have the following properties of the generator polynomial.

PROPOSITION 1.44. Let $g_C$ be the generator polynomial for a cyclic code $C$. Then

1. $g_C$ divides $t^n - 1$ in $\mathbb{F}_q[t]$;
2. $\dim C = n - \deg g_C$. 

PROOF. (1) Divide \( t^n - 1 \) by \( g_C \) with remainder: \( t^n - 1 = hg_C + r \) for some \( h, r \in \mathbb{F}_q[t] \), where either \( r = 0 \) or \( \deg r < \deg g_C \). This implies that in \( R = \mathbb{F}_q[t]/(t^n - 1) \) we have \( 0 = hg_C + r \), and so \( r = -hg_C \in I_C \). But \( g_C \), being a generator for \( I_C \), is a polynomial of smallest positive degree in \( I_C \). Thus, \( r = 0 \) in \( R \), and also in \( \mathbb{F}_q[t] \). Therefore, \( t^n - 1 = hg_C \) in \( \mathbb{F}_q[t] \), i.e. \( g_C \) divides \( t^n - 1 \) in \( \mathbb{F}_q[t] \).

(2) Let \( l = \deg g_C \). According to our correspondence \( c \mapsto c(t) \), the code \( C \) and the ideal \( I_C \) are isomorphic as vector spaces over \( \mathbb{F}_q \). We have \( c(t) \in I_C = \langle g_C \rangle \), so \( c(t) = h(t)g_C(t) \) for some \( h \) with \( \deg h \leq n - 1 - l \). This implies that \( I_C \), as a vector space over \( \mathbb{F}_q \), has a basis \( \{g_C, tg_C, \ldots, t^{n-1}g_C\} \). Therefore, \( \dim C = \dim I_C = n - l \).

EXAMPLE 1.45. Let \( C_{m,\mathbb{F}_q^*} \) be the Reed–Solomon code with \( \mathcal{P} = \mathbb{F}_q^* = \mathbb{F}_q \setminus \{0\} \). In Exercise 1.4 you will show a code is cyclic if and only of the cyclic permutation of every row of a generator matrix of \( C \) lies in \( C \). We computed a generator matrix \( G \) for the Reed–Solomon code in (1.1). In our case \( \mathcal{P} = \mathbb{F}_q^* \), which is known to be a cyclic group under multiplication. Hence, we can write all the elements of \( \mathbb{F}_q^* \) as powers of a single element: \( \mathbb{F}_q^* = \{1, \alpha, \ldots, \alpha^{q-2}\} \). The generator matrix \( G \) then becomes:

\[
G = \begin{bmatrix}
1 & 1 & \ldots & 1 \\
1 & \alpha & \ldots & \alpha^{q-2} \\
1 & \alpha^2 & \ldots & \alpha^{2(q-2)} \\
\vdots & \vdots & \ddots & \vdots \\
1 & \alpha^m & \ldots & \alpha^{m(q-2)}
\end{bmatrix}.
\]

Check that the cyclic permutation of each row of \( G \) is a constant multiple of this row, and hence lies in \( C_{m,\mathbb{F}_q^*} \). This shows that \( C_{m,\mathbb{F}_q^*} \) is a cyclic code.

It is an interesting exercise to compute the generator polynomial for the code \( C_{m,\mathbb{F}_q^*} \). We outline the steps in Exercise 1.7.

1.5. Asymptotic of Codes

In this section we will look at the asymptotic behavior of parameters of codes with increasing length. Recall that for an \([n, k, d]_{\mathbb{F}_q}\)-code, the quantities \( R = k/n \) and \( \delta = d/n \) are called the information rate and the relative minimum distance, respectively. It is immediate that \( 0 \leq \delta \leq 1 \) and \( 0 \leq R \leq 1 \). Therefore we may consider points in the unit square with coordinates \((\delta, R)\) which correspond to the relative parameters of codes. Here is the main problem, somewhat loosely stated.

PROBLEM. Give a description of the set

\[ V_q = \{(\delta, R) \in [0, 1]^2 \mid \text{there exists an } [n, k, d]_{\mathbb{F}_q}\text{-code with } \delta = d/n, R = k/n\} \].

The following theorem by Shannon claims the existence of codes with very good parameters if we allow the length of the code to be large. We will state the theorem without giving much detail, hoping that the result can still be appreciated.

First, let \( p \) denote the probability with which an error occurs in a codeword. There is an explicit function of \( p \), called the capacity of the channel of transmission, given by

\[ \text{capacity}(p) = 1 + p \log_q p + (1 - p) \log_q(1 - p) \].
Theorem 1.46. (Shannon’s Theorem) For any \( \epsilon > 0 \), any \( 0 < p < 1 \), and any \( R < \text{capacity}(p) \) there exists a code with information rate at least \( R \), probability of error per codeword \( p \), and probability of incorrect decoding less than \( \epsilon \).

In fact, \( n \to \infty \), as \( \epsilon \to 0 \), so to achieve high reliability we need to consider long codes.

Next consider the set of limit points of \( V_q \):

\[
U_q = \{ (\delta, R) \in [0,1]^2 \mid \text{there exists a sequence of distinct } [n_i, k_i, d_i]_q \text{-codes with } \delta = \lim_{i \to \infty} d_i/n_i, R = \lim_{i \to \infty} k_i/n_i \}.
\]

We remark that \( n_i \to \infty \), as \( i \to \infty \), since there are only finitely many distinct codes of bounded length.

Definition 1.47. We say an infinite family of distinct \( [n_i, k_i, d_i]_q \) codes is asymptotically good if the corresponding \( \delta = \lim_{i \to \infty} d_i/n_i \) and \( R = \lim_{i \to \infty} k_i/n_i \) are both positive.

Towards a solution to the main problem, Yu. Manin gave the following description of the set \( U_q \):

Theorem 1.48. There exists a continuous function \( \alpha_q(\delta) \) such that

\[
U_q = \{ (\delta, R) \in [0,1]^2 \mid 0 \leq \delta \leq 1 \text{ and } 0 \leq R \leq \alpha_q(\delta) \}.
\]

Moreover, \( \alpha_q \) decreases on the segment \( [0, \frac{2-1}{q}] \) and vanishes on \( [\frac{2-1}{q}, 1] \).

Many questions about this function remain open, e.g. it is unknown whether it is differentiable or concave up.

If we restrict ourselves to only linear codes we can similarly define the sets \( V_q^{lin} \) and \( U_q^{lin} \). In fact, Manin’s theorem guarantees existence of the corresponding function \( \alpha_q^{lin} \), as well. Clearly, \( \alpha_q(\delta)^{lin} \leq \alpha_q(\delta) \) for any \( 0 \leq \delta \leq 1 \), but it is not known whether they are indeed the same function.

The functions \( \alpha_q \) and \( \alpha_q^{lin} \) have been intensively studied by many mathematicians. The main goal was to write down non-trivial (and hopefully tight) upper and lower bounds for these functions. In the next subsections we will only present some of the non-trivial bounds.

1.5.1. Upper Bounds. Our first theorem, called the Singleton bound, is a general result relating the three parameters of a code.

Theorem 1.49. (The Singleton bound) Let \( C \) be an \( [n, k, d]_q \)-code. Then

\[
d \leq n - k + 1.
\]

Proof. Let \( M \) be the number of elements of \( C \). The key observation is that after erasing the last \( d - 1 \) entries in every codeword in \( C \) we obtain \( M \) distinct words of length \( n - d + 1 \). Indeed, if two of the them were the same that would mean that the original codewords had at least \( n - d + 1 \) common entries, and their Hamming distance would be less than \( d \), which is impossible since \( d \) is the minimum distance of \( C \). This implies that \( M \leq q^{n-d+1} \), which is the total number of words of length \( n - d + 1 \). Taking the logarithm we obtain \( k = \log_q(M) \leq n - d + 1 \), as stated.
Codes that meet the Singleton bound are called maximum distance separable or, simply, MDS codes. For example, any Reed–Solomon code $C_{m,p}$ is an MDS code. Indeed, by Theorem 1.28, $C_{m,p}$ is an $[n, m+1, n-m]_q$-code and we see that $d = n - m = n - (m + 1) + 1 = n - k + 1$.

**Corollary 1.50.** (Asymptotic Singleton bound)

$$\alpha_q(\delta) \leq 1 - \delta.$$  

**Proof.** By Theorem 1.48, we need to show that for any $(\delta, R) \in U_q$ we have $R \leq 1 - \delta$. Indeed, given a sequence of distinct $[n_i, k_i, d_i]_q$-codes with $\delta = \lim_{i \to \infty} d_i/n_i$ and $R = \lim_{i \to \infty} k_i/n_i$, apply the Singleton bound for every code in the sequence: $k_i \leq n_i - d_i + 1$. Now dividing by $n_i$ and taking the limit as $i \to \infty$ we obtain the required inequality $R \leq 1 - \delta$. \(\square\)

Next we will prove a different upper bound for $\alpha_q$, called the Asymptotic Hamming bound. We will start with the Hamming bound for a single code.

**Theorem 1.51.** (The Hamming bound) For any $[n, k, d]_q$-code we have

$$\sum_{i=0}^{\left\lfloor \frac{d-1}{2} \right\rfloor} \binom{n}{i} (q-1)^i \leq q^{n-k}.$$

**Proof.** Recall the formula for the number of elements in a ball in Section 1.1.4. Here we take a ball of radius $\left\lfloor \frac{d-1}{2} \right\rfloor$ around each codeword. Since the balls are disjoint, their union contains words of length $n$. This cannot exceed the total number of words of length $n$, which is $q^n$. Therefore,

$$|C| \sum_{i=0}^{\left\lfloor \frac{d-1}{2} \right\rfloor} \binom{n}{i} (q-1)^i \leq q^n.$$

It remains to note that $|C| = q^k$, and the required inequality follows. \(\square\)

**Definition 1.52.** The following is the Hamming entropy function

$$H_q(t) = \begin{cases} t \log_q(q-1) - t \log_q(t) - (1-t) \log_q(1-t), & \text{if } 0 < t \leq \frac{q-1}{q}, \\ 0, & \text{if } t = 0 \end{cases}.$$

**Lemma 1.53.**

$$H_q\left(\frac{t}{n}\right) \sim \frac{1}{n} \log_q \left(\sum_{i=0}^{t} \binom{n}{i} (q-1)^i\right), \quad \text{as } n \to \infty.$$

**Proof.** The proof uses Stirling’s formula and is left as an exercise. \(\square\)

**Corollary 1.54.** (Asymptotic Hamming bound)

$$\alpha_q(\delta) \leq 1 - H_q\left(\frac{\delta}{2}\right).$$
1.5. ASYMPTOTIC OF CODES

Proof. As in the proof of Corollary 1.50, it is enough to show that for any $(\delta, R) \in U_q$ we have $R \leq 1 - H_q(\delta/2)$. Again, consider a sequence of distinct $[n_i, k_i, d_i]_q$-codes with $\delta = \lim_{i \to \infty} d_i/n_i$ and $R = \lim_{i \to \infty} k_i/n_i$, and apply the Hamming bound (Theorem 1.51) to each of the code in the sequence. Taking the log of both sides of and dividing by $n$ we obtain

$$\frac{1}{n} \log \left( \sum_{i=0}^{\lfloor \frac{d_i}{2} \rfloor} \binom{n_i}{i} (q - 1)^i \right) \leq 1 - \frac{k_i}{n_i}.$$ 

It remains to take the limit as $i \to \infty$ and apply Lemma 1.53.

Below we depict both the asymptotic Singleton (in red) and the asymptotic Hamming (in green) bounds for $q = 8$. Note that the latter intersects the horizontal axis at $\delta = 2(q - 1)/q$.

![Figure 1.1. The asymptotic Singleton and Hamming bounds for $q = 8$.](image)

1.5.2. Lower Bounds. The question of finding a lower bound for the function $\alpha_q(\delta)$ amounts to finding points $(\delta, R)$ in $U_q$ with largest possible value of $R$. This, in turn, concerns the existence of codes of size $q^k$, given the values of $n$ and $d$. Of course, if one such code exists then by discarding some of its codewords with can obtain a smaller code with the same $n$ and $d$. So it make sense to consider codes of largest possible size with given $n$ and $d$. This motivates the following definition.

**Definition 1.55.** Define $A_q(n, k)$ to be the size of the largest code of length $n$ and minimum distance $d$, i.e.

$$A_q(n, d) = \max \{ q^k \mid \text{there exists an } [n, k, d]_q\text{-code} \}.$$ 

The following result is due to Gilbert.
Theorem 1.56. (The Gilbert Bound)

\[ A_q(n, d) \geq \frac{q^n}{\sum_{i=0}^{d-1} \binom{n}{i}(q-1)^i}. \]

Proof. Let \( C \) be a \([n, k, d]_q\)-code of size \(|C| = A_q(n, d)\). This means we cannot find a word in \( A^n \setminus C \) whose distance from every codeword in \( C \) is greater than or equal to \( d \) (otherwise we would have included it in \( C \)). Therefore the union of the balls of radius \( d - 1 \) centered at the codewords of \( C \) must contain all the words of length \( n \). This implies

\[ |C| \left( \sum_{i=0}^{d-1} \binom{n}{i}(q-1)^i \right) \geq q^n, \]

and the bound follows. \( \Box \)

A corollary from this is the asymptotic Gilbert bound.

Corollary 1.57. (Asymptotic Gilbert bound)

\[ \alpha_q(\delta) \geq 1 - H_q(\delta). \]

Proof. Consider a sequence of \([n_i, k_i, d_i]_q\)-codes with \( \delta = \lim_{i \to \infty} d_i/n_i \) and \( R = \lim_{i \to \infty} k_i/n_i \), and such that \( k_i = A_q(n_i, d_i) \) for every \( i \geq 1 \). By Theorem 1.48 \( R \leq \alpha_q(\delta) \). On the other hand, by the Gilbert bound

\[ k_i = \log_q(A_q(n_i, d_i)) \geq n_i - \log_q \left( \sum_{i=0}^{d-1} \binom{n_i}{i}(q-1)^i \right). \]

Dividing both sides by \( n_i \) and taking the limit as \( i \to \infty \), with the help of Lemma 1.53 we obtain

\[ R \geq 1 - H_q(\delta). \]

Remarkably, the same lower bound holds for \( \alpha_{q}^{lin}(\delta) \) as well. This is not trivial, since it requires existence of linear codes with the above properties. This result, which we will not prove, is called the Gilbert–Varshamov bound.

Theorem 1.58. (Gilbert–Varshamov bound)

\[ \alpha_{q}^{lin}(\delta) \geq 1 - H_q(\delta). \]

Much more is known about this bound, for example, that almost all linear codes (i.e. points of \( V_q^{lin} \)) lie on the graph of \( R_{GV} = 1 - H_q(\delta) \). It was believed for a while that the Gilbert–Varshamov bound cannot be improved until Tsfasman, Vlăduț, and Zink found a bound which beats the Gilbert–Varshamov bound on a certain segment for large enough \( q \). To achieve that they studied codes arising from algebraic curves over finite fields, which now are commonly called algebraic geometry codes (AG codes). We will study them in Chapter 5. Right now we state the Tsfasman–Vlăduț–Zink bound.

Theorem 1.59. Let \( q \) be an even power of a prime. Then

\[ \alpha_{q}^{lin}(\delta) \geq 1 - \delta - \frac{1}{\sqrt{q}-1}. \]
One can check that for \( q \geq 49 \) the graph of \( R_{TVZ} = 1 - \frac{1}{\sqrt{q-1}} \) intersects the graph of \( R_{GV} = 1 - H_q(\delta) \) on the segment \([\delta_1, \delta_2]\), where \( \delta_1 \) and \( \delta_2 \) are the roots of the equation \( H_q(\delta) - \delta = \frac{1}{\sqrt{q-1}} \) (see Exercise 1.12).

We finish this chapter with a graph of the upper and lower bounds we discussed in this section. For \( q = 81 \) we plotted the Singleton bound (red), the Hamming bound (green), the Gilbert–Varshamov bound (yellow), and the Tsfasman–Vlăduţ–Zink bound (blue).

\[ \text{Figure 1.2. Upper bounds: the Singleton and Hamming bounds; lower bounds: the Gilbert–Varshamov, and the Tsfasman–Vlăduţ–Zink bounds for } q = 81. \]

**Exercises**

**Exercise 1.1.** Prove that \( d = 3 \) for the 4-7 Hamming Code. (Hint: show that the weight of each codeword is at least 3.)

**Exercise 1.2.** Show that the Hamming distance \( d \) defines a metric on the set of all words of length \( n \) in the alphabet \( \mathcal{A} \), i.e. for any \( \mathbf{a}, \mathbf{b}, \) and \( \mathbf{c} \) in \( \mathcal{A}^n \) we have

1. \( d(\mathbf{a}, \mathbf{b}) \geq 0 \),
2. \( d(\mathbf{a}, \mathbf{b}) = 0 \) if and only if \( \mathbf{a} = \mathbf{b} \),
(3) \( d(a, b) = d(b, a) \),
(4) \( d(a, c) \leq d(a, b) + d(b, c) \).

Exercise 1.3. Write the generator matrix for the dual to the Reed-Solomon code in Example 1.30 using the residue map. Find the parameters of this code.

Exercise 1.4. Let \( G \) be a generator matrix for a linear code \( C \). Show that \( C \) is cyclic if and only if the cyclic permutation of every row of \( G \) lies in \( C \).

Exercise 1.5. Let \( G \) be the (standard) generating matrix for \( C_{m, F^n_q} \). Show that \( C \) is cyclic if and only if the cyclic permutation of every row of \( G \) is a multiple of this row. Deduce that the Reed-Solomon code \( C_{m, F^n_q} \) is cyclic. (Hint: Use the fact that \( F^n_q \) is a cyclic group under multiplication; apply Exercise 1.4.)

Exercise 1.6. Let \( \beta \) be an element of \( F_q \), \( \beta \neq 1 \). Show that \( \sum_{j=0}^{q-2} \beta^j = 0 \). (Hint: Use the fact that \( F^n_q \) is a cyclic group under multiplication, and hence the elements of \( F_q \) are the roots of \( t^q - t \).)

Exercise 1.7. In this exercise you will find the generator polynomial \( g_C(t) \) for the Reed-Solomon code \( C = C_{m, F^n_q} \). We have \( F^n_q = \langle \alpha \rangle \) for some \( \alpha \in F_q^* \). By Proposition 1.44, \( g_C \) divides \( t^q - 1 \), hence, \( g_C(t) = \prod_{\beta \in S} (t - \beta) \) for some subset \( S \subset F_q^* \) (see the hint in Exercise 1.6). Also from Proposition 1.44 we see that \( |S| = q - m - 2 \).

(a) Show that \( S \) is contained in the set of common roots of all polynomials \( c(t) \in I_C \).
(b) Let \( c = (f(1), f(\alpha), \ldots, f(\alpha^{q-2})) \) be a codeword in \( C \), corresponding to a polynomial \( f \in L(m) \). Write down \( c(t) \). Show that \( c(t) \) vanishes at \( t = \alpha^k \) for \( 1 \leq k \leq q - m - 2 \). (Hint: Use Exercise 1.6.)
(c) Use parts (a) and (b) to find \( g_C(t) \).

Exercise 1.8. Let \( C \) be an \([n, k, d]_q\)-code. Show that the minimum distance of the dual code \( C^\perp \) cannot be greater than \( k + 1 \).

Exercise 1.9. Recall the construction of the field of four elements: \( F_4 = \{0, 1, \alpha, 1 + \alpha\} \), where addition is mod 2 and \( \alpha \) is an element satisfying \( \alpha^2 = 1 + \alpha \). Show that the linear code with generator matrix

\[
\begin{bmatrix}
1 & 0 & 1 & 1 \\
0 & 1 & \alpha & \alpha^2
\end{bmatrix}
\]

is an MDS-code.

Exercise 1.10. Compute a generator matrix for the dual to the code in Exercise 1.9 and show it is also an MDS-code.

Exercise 1.11. Let \( C \) be an MDS-code with parity-check matrix \( H \). Prove that any \( n - k \) of the rows of \( H \) are linearly independent. (Hint: Assuming the opposite show that \( C \) contains a codeword with weight less than or equal to \( n - k \).)
Exercise 1.12. Let $H_q$ be the Hamming entropy function. Show that for $q \geq 49$ the line $R = 1 - \delta - \frac{1}{\sqrt{q-1}}$ intersects the graph of the function $R = 1 - H_q(\delta)$ on the segment $[\delta_1, \delta_2]$, where $\delta_1$ and $\delta_2$ are the roots of the equation

$$H_q(\delta) - \delta = \frac{1}{\sqrt{q-1}}.$$  

(Hint: Find the maximum of $F(\delta) = H_q(\delta) - \delta$ on $[0, (q-1)/q]$. Check that the maximum is attained at $\delta = \frac{q-1}{2\sqrt{q-1}}$ and equals $\log_q(2q-1) - 1$. Then show that this maximum is greater than $\frac{1}{\sqrt{q-1}}$ for $q \geq 49$ by plotting both as functions of $q$. You can use Maple or any other computer system.)

Exercise 1.13. Prove that if $C$ is an MDS-code then so is its dual $C^\perp$. (Hint: Use Exercise 1.8 and Exercise 1.11.)
CHAPTER 2

Algebraic Curves

2.1. Fields and Polynomial Rings

2.1.1. Finite Fields. You are already familiar with a field of prime order, 
\( \mathbb{F}_p = \{0, 1, \ldots, p - 1\} \). It consists of classes of integers modulo a prime \( p \). There are other examples of fields, e.g. we have already seen the field of four elements 
\( \mathbb{F}_4 = \{0, 1, \alpha, 1 + \alpha\} \) in Exercise 1.9. Recall that we use the identities 
\( \alpha^2 = 1 + \alpha \) and \( 1 + 1 = 0 \). Here are the addition and the multiplication tables for 
\( \mathbb{F}_4 \).

\[
\begin{array}{c|cccc}
+ & 0 & 1 & \alpha & 1 + \alpha \\
0 & 0 & 1 & \alpha & 1 + \alpha \\
1 & 1 & 0 & 1 + \alpha & \alpha \\
\alpha & \alpha & 1 + \alpha & 0 & 1 \\
1 + \alpha & 1 + \alpha & \alpha & 1 & 0 \\
\end{array}
\]

\[
\begin{array}{c|cccc}
\cdot & 0 & 1 & \alpha & 1 + \alpha \\
0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & \alpha & 1 + \alpha \\
\alpha & 1 + \alpha & 0 & \alpha & 1 + \alpha \\
1 + \alpha & \alpha & 1 + \alpha & 1 & 0 \\
\end{array}
\]

In fact, we can define it as the quotient ring 
\( \mathbb{F}_4 = \mathbb{F}_2[x] / (x^2 + x + 1) \). Let 
\( h = x^2 + x + 1 \in \mathbb{F}_2[x] \). Since \( h \) has degree 2, the possible remainders mod \( h \) are 
either constants in \( \mathbb{F}_2 \) or linear functions over \( \mathbb{F}_2 \). Thus, we can write 
\( \mathbb{F}_4 = \{a_0 + a_1 x \mid a_i \in \mathbb{F}_2\} = \{0, 1, x, 1 + x\} \), 
assuming that these are classes mod \( h \). Note that \( x^2 \equiv x + 1 \mod h \), so in our 
notation before \( \alpha \) denotes the class of \( x \mod h \). You can check that the addition 
and multiplication on classes mod \( h \) is exactly the one described in the above tables. 
For example, \( (1 + x)x = x + x^2 \equiv 1 \mod h \), hence \( (1 + \alpha)\alpha = 1 \).

Let \( \mathbb{F} \) be a field. Recall that a polynomial \( h \in \mathbb{F}[x] \) is called irreducible over \( \mathbb{F} \) 
if \( h \) cannot be written as a product of two polynomials in \( \mathbb{F}[x] \) of positive degree. 
Here is a standard fact from abstract algebra.

**Proposition 2.1.** Let \( \mathbb{F} \) be a field and \( h \in \mathbb{F}[x] \). The quotient ring \( \mathbb{F}[x] / \langle h \rangle \) is 
a field if and only if \( h \) is irreducible over \( \mathbb{F} \).

You should check that \( x^2 + x + 1 \) is irreducible over \( \mathbb{F}_2 \), and so our quotient 
ring \( \mathbb{F}_2[x] / \langle x^2 + x + 1 \rangle \) is indeed a field.

This construction can be generalized to produce fields of size \( p^n \) for any prime \( p \) 
and any integer \( n \geq 1 \). Let \( h \) be an irreducible polynomial over \( \mathbb{F}_p \). Then 
\( \mathbb{F}_{p^n} = \mathbb{F}_p[x] / \langle h \rangle \) is a field of \( p^n \) elements. We have 
\( \mathbb{F}_{p^n} = \{a_0 + a_1 \alpha + \cdots + a_{n-1} \alpha^{n-1} \mid a_i \in \mathbb{F}_p\} \), 
where again \( \alpha \) is the classes of \( x \mod h \). Since there are exactly \( p \) choices for every 
coefficient \( a_i \) we obtain \( p^n \) such distinct classes. By Proposition 2.1 this is a field 
since we assumed \( h \) to be irreducible. The question now, of course, is “Do there 
exist irreducible polynomials over \( \mathbb{F}_p \) for any prime \( p \) of any given degree \( n \)?” The
answer is “yes”, and it can be shown by just counting the number of reducible polynomials of degree $n$ (i.e. products of polynomials of smaller positive degrees) and checking that it is less than the total number of polynomials of degree $n$ over $\mathbb{F}_p$.

We will not be doing this here, but if you are interested you can find it in [?].

Recall that if $\mathbb{F}$ and $\mathbb{K}$ are two fields with the same operation, the same 0,1 elements, and such that $\mathbb{F} \subset \mathbb{K}$, we say that $\mathbb{F} \subset \mathbb{K}$ is a field extension. For example $\mathbb{F}_2 \subset \mathbb{F}_4$ is a field extension. Note that in this case $\mathbb{K}$ is a vector space over $\mathbb{F}$. An field extension $\mathbb{F} \subset \mathbb{K}$ is called finite if $\mathbb{K}$ is a finite dimensional space over $\mathbb{F}$. In particular, if $\mathbb{K}$ is a finite field then $\mathbb{F} \subset \mathbb{K}$ is a finite field extension. In this case $\mathbb{K}$ must have a basis $\{v_1, \ldots, v_k\} \subset \mathbb{K}$ over $\mathbb{F}$, i.e.

$$\mathbb{K} = \{c_1v_1 + \cdots + c_kv_k \mid c_i \in \mathbb{F}\},$$

$k = \dim_{\mathbb{F}} \mathbb{K}$. For example, $\mathbb{F}_4$ is a 2-dimensional vector space over $\mathbb{F}_2$ with a basis $\{1, \alpha\}$. More generally, $\mathbb{F}_{p^n}$ is an $n$-dimensional vector space over $\mathbb{F}_p$ with a basis $\{1, \alpha, \ldots, \alpha^{n-1}\}$.

**Proposition 2.2.** Let $\mathbb{F} \subset \mathbb{K} \subset \mathbb{L}$ be a “tower” of field extensions and $\mathbb{L}$ is finite dimensional over $\mathbb{F}$. Then

$$\dim_{\mathbb{F}} \mathbb{L} = \dim_{\mathbb{F}} \mathbb{K} \dim_{\mathbb{K}} \mathbb{L}.$$  

**Proof.** Start with a basis $\{w_1, \ldots, w_l\}$ for $\mathbb{L}$ over $\mathbb{K}$ and a basis $\{v_1, \ldots, v_k\}$ for $\mathbb{K}$ over $\mathbb{F}$ and show that the set of pairwise products $\{v_iw_j \mid 1 \leq i \leq k, 1 \leq j \leq l\}$ forms a basis for $\mathbb{K}$ over $\mathbb{F}$. You are invited to fill the details yourself. □

Here is the first main result in the theory of finite fields.

**Theorem 2.3.** Let $\mathbb{F}$ be a finite field. Then

1. $\mathbb{F}$ has $p^n$ elements for some $p$ and $n \geq 1$.
2. $\mathbb{F}$ is isomorphic to $\mathbb{F}_p[x]/\langle h \rangle$ for some monic irreducible degree $n$ polynomial $h$ in $\mathbb{F}_p[x]$.

**Proof.** Since $\mathbb{F}$ is finite, the elements $1, 1+1, 1+1+1, \ldots$ cannot all be distinct. Therefore, there exists the smallest integer $p \geq 2$ such that $1 + \cdots + 1 = 0$. Note that $p$ must be prime, otherwise if $p = rs$ then

$$0 = 1 + \cdots + 1 = \underbrace{(1 + \cdots + 1)}_{p \text{ times}} \underbrace{(1 + \cdots + 1)}_{r \text{ times}} \underbrace{(1 + \cdots + 1)}_{s \text{ times}},$$

which means that $\mathbb{F}$ has zero divisors, which is impossible since $\mathbb{F}$ is a field. (The elements $1 + \cdots + 1$ and $1 + \cdots + 1$ are non-zero by the minimality of $p$.) Such $p$ is called the characteristic of the field $\mathbb{F}$, $p = \text{char} \mathbb{F}$. This follows that $\mathbb{F}$ contains $\mathbb{F}_p$ as a subfield.

1. We have a field extension $\mathbb{F}_p \subset \mathbb{F}$, so by above $\mathbb{F}$ is a finite dimensional vector space over $\mathbb{F}_p$, i.e.

$$\mathbb{F} = \{c_1v_1 + \cdots + c_nv_n \mid c_i \in \mathbb{F}_p\},$$

for a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{F}$ over $\mathbb{F}_p$. Therefore $\mathbb{F}$ has $p^n$ elements.

2. Let $q = p^n$. We have already mentioned that $\mathbb{F}^* = \mathbb{F} \setminus \{0\}$ is a cyclic group under multiplication (for a proof see, for example [?]), i.e. $\mathbb{F}^* = \{1, \alpha, \ldots, \alpha^{q-2}\}$ for some $\alpha \in \mathbb{F}^*$.Therefore $\alpha$ satisfies $\alpha^{q-1} = 1$ by the Lagrange theorem. In other
words, \( \alpha \) is a root of the polynomial \( x^{q-1} - 1 \) in \( \mathbb{F}_p[x] \). Let \( h \) be the monic irreducible factor of \( x^{q-1} - 1 \), for which \( \alpha \) is a root. Then \( \mathbb{F} \) is isomorphic to \( \mathbb{F}_p[x]/\langle h \rangle \). \( \square \)

**Definition 2.4.** An element \( \alpha \) in \( \mathbb{F}_q \) is called *primitive* if it generates the multiplicative group \( \mathbb{F}_q^* \).

As it follows from group theory, a cyclic group of order \( k \) has \( \varphi(k) \) generators, where \( \varphi \) is the Euler function. Therefore, in every field of order \( q \) there are \( \varphi(q-1) \) primitive elements.

**Example 2.5.** There are \( \varphi(3) = 2 \) primitive elements in \( \mathbb{F}_4 \), namely \( \alpha \) and \( 1 + \alpha \).

Now we will state the complete classification of finite fields.

**Theorem 2.6.**
1. For any prime \( p \) and any \( n \geq 1 \) there exists a finite field of order \( p^n \).
2. Any two fields of the same size are isomorphic.

**Proof.** Part (1) follows from our discussion after Proposition 2.1. For part (2) we need something stronger than what we used in the proof of Theorem 2.3. We saw that \( \alpha \) is a root of \( x^{p^n-1} - 1 \). This implies that every element \( \alpha^i \) of \( \mathbb{F}^* \) is a root of \( x^{p^n-1} - 1 \), and hence, \( \mathbb{F} \) consists of the roots of \( x^p - x \). In this case we say that \( \mathbb{F} \) is the splitting field of \( x^p - x \). Now we need a theorem from field theory which says that the splitting field of a polynomial is unique up to an isomorphism. \( \square \)

Here is another example of a finite field.

**Example 2.7.** To construct a field of order 9 as a quotient ring \( \mathbb{F}_9 = \mathbb{F}_3[x]/\langle h \rangle \) we need a degree 2 irreducible polynomial \( h \) over \( \mathbb{F}_3 \). For example, one can take \( h(x) = 1 + x^2 \). Let \( \alpha \) be the class of \( x \mod h \). Then \( \alpha \) satisfies \( 1 + \alpha^2 = 0 \). We obtain

\[
\mathbb{F}_9 = \{ a_0 + a_1 \alpha \mid a_i \in \mathbb{F}_3 \} = \{ 0, 1, 2, \alpha, 2\alpha, 1 + \alpha, 2 + \alpha, 1 + 2\alpha, 2 + 2\alpha \}.
\]

Let us compute some products and sums in \( \mathbb{F}_9 \).

- \((1 + \alpha)(2 + \alpha) = 2 + 3\alpha + \alpha^2 = 2 + 0 + (-1) = 1\); hence \((1 + \alpha)^{-1} = 2 + \alpha\)
- \((1 + 2\alpha)^2 = 1 + 4\alpha + 4\alpha^2 = 1 + \alpha + (-1) = \alpha\)
- \(\alpha + (2 + 2\alpha) = 2 + 3\alpha = 2\)

Let \( \mathbb{F}_q \) be a finite field of order \( q = p^n \), where \( p \) is the characteristic of the field. We will define a very important map from \( \mathbb{F}_q \) to itself.

**Definition 2.8.** Let \( \mathbb{F}_q \) be a finite field of characteristic \( p \). The map

\[
\sigma : \mathbb{F}_q \to \mathbb{F}_q, \quad \alpha \mapsto \alpha^p
\]

is called the *Frobenius automorphism*.

Here some of its properties.

**Proposition 2.9.** Let \( \mathbb{F}_q \) be a finite field of \( q = p^n \) elements. Then
1. for any \( \alpha, \beta \in \mathbb{F}_q \) we have \((\alpha + \beta)^p = \alpha^p + \beta^p \);
2. the map \( \alpha \mapsto \alpha^p \) is an automorphism of \( \mathbb{F}_q \) which fixes \( \mathbb{F}_p \);
3. the Galois group of all automorphisms of \( \mathbb{F}_q \) which fix \( \mathbb{F}_p \),

\[
\text{Gal}(\mathbb{F}_q) = \{ \phi : \mathbb{F}_q \to \mathbb{F}_q \mid \phi(a) = a, \forall a \in \mathbb{F}_p \},
\]

is cyclic of order \( n \), generated by \( \sigma \).
Proof. (1) By the binomial formula

\[(\alpha + \beta)^p = \sum_{i=0}^{p} \binom{p}{i} \alpha^{p-i} \beta^i = \alpha^p + \sum_{i=0}^{p-1} \binom{p}{i} \alpha^{p-i} \beta^i = 0 + \beta^p = \alpha^p + \beta^p,\]

where the middle terms are all zero since \(p\) divides \(\binom{p}{i}\) for \(1 \leq i \leq p - 1\) and \(p\) is the characteristic of the field.

(2) By part (1), \(\sigma(\alpha + \beta) = \sigma(\alpha) + \sigma(\beta)\). Also \(\sigma(\alpha \beta) = (\alpha \beta)^p = \sigma(\alpha) \sigma(\beta)\), hence, \(\sigma\) is a ring homomorphism. Next, \(\text{Ker}(\sigma) = \{\alpha \in \mathbb{F}_q \mid \alpha^p = 0\} = \{0\}\), i.e. \(\sigma\) is injective. But \(\mathbb{F}_q\) is finite, so any injective map is also surjective. Therefore \(\sigma\) is an automorphism of \(\mathbb{F}_q\). The fact that \(\sigma\) fixes any \(a \in \mathbb{F}_p\) is the Fermat Little Theorem: \(\sigma(a) = a^p = a\) for any \(a \in \mathbb{F}_p\).

(3) From Galois theory we know that \(\mathbb{F}_p \subset \mathbb{F}_q\) is a Galois extension, so \(|\text{Gal}(\mathbb{F}_q)| = \dim_{\mathbb{F}_p} \mathbb{F}_q = n\). Let us show that the subgroup generated by \(\sigma\) has \(n\) distinct elements:

\[\langle \sigma \rangle = \{\text{id}, \sigma, \sigma^2, \ldots, \sigma^{n-1}\}.\]

Here \(\sigma^i\) is the composition of \(\sigma\) with itself \(i\) times: \(\sigma^i = \sigma \circ \cdots \circ \sigma\) (\(i\) times), so \(\sigma^i(\alpha) = \alpha^{p^i}\).

Indeed, if \(\sigma^i = \sigma^j\) then \(\alpha^{p^i} = \alpha^{p^j}\) for any \(\alpha \in \mathbb{F}_q\), in particular, when \(\alpha\) is a primitive element. Then \(\alpha^{p^i-p^j} = 1\), which implies that \(\text{ord}(\alpha) = p^n - 1\) divides \(p^i - p^j\). Since both \(i, j\) are less than \(n\) this is only possible when \(i = j\).

The next result describes all possible subfields of \(\mathbb{F}_q\).

**Theorem 2.10.** Let \(K\) be a subfield of \(\mathbb{F}_q\), where \(q = p^n\). Then \(K\) is isomorphic to \(\mathbb{F}_{p^k}\) for some divisor \(k\) of \(n\). Moreover,

\[K = \{\beta \in \mathbb{F}_q \mid \sigma^k(\beta) = \beta\} .\]

Proof. Clearly, \(K\) has the same characteristic \(p\), so \(|K| = p^k\) for some \(k\). To see that \(k\) must be a divisor of \(n\) consider a tower \(\mathbb{F}_p \subset K \subset \mathbb{F}_q\). By Proposition 2.2

\[n = \dim_{\mathbb{F}_p} \mathbb{F}_q = \dim_{\mathbb{F}_p} K \dim_{\mathbb{F}_q} \mathbb{F}_q = k \dim_{\mathbb{F}_q} \mathbb{F}_q ,\]

hence \(k | n\). The second statement follows from the Fundamental Theorem of the Galois theory and will not be proved here.

**Example 2.11.** Let us describe all subfields of \(\mathbb{F}_{2^6}\). There are four divisors of \(6\) including 1 and 6. They correspond to the four subfields:

\[
\begin{array}{c}
\mathbb{F}_{2^2} \\
\mathbb{F}_{2} \\
\mathbb{F}_{2^3} \\
\mathbb{F}_{2^6}
\end{array}
\]

\[
\begin{array}{c}
\mathbb{F}_{2^2} \\
\mathbb{F}_{2} \\
\mathbb{F}_{2^3} \\
\mathbb{F}_{2^6}
\end{array}
\]
Example 2.12. We can construct an infinite sequence of subfields of characteristic 2:
\[ F_2 \subset F_{2^2} \subset F_{2^3} \subset \cdots \subset F_{2^{(n-1)!}} \subset F_{2^n} \subset \ldots \]
Note that \((n-1)!\) divides \(n!\) for any \(n \geq 1\), so these are indeed subfields.

2.1.2. Algebraic Closure. Remember the Fundamental Theorem of Algebra which says that any polynomial of degree \(n\) over complex numbers has exactly \(n\) complex roots, counting with multiplicities. This is the property of the complex numbers begin algebraically closed. What are algebraically closed fields of positive characteristic?

We will start with the definition.

Definition 2.13. A field \(F\) is called algebraically closed if every polynomial \(f \in F[x]\) has a root in \(F\).

For example, \(\mathbb{R}\) is not algebraically closed, since \(1 + x^2 \in \mathbb{R}[x]\), but has no real roots; \(\mathbb{C}\) is algebraically closed as we mentioned before. Notice that this implies that \(f\) has all of its roots in \(F\), and so \(f\) splits into a product of linear factors in \(F[x]\).

Definition 2.14. Let \(F\) be a field. The algebraic closure of \(F\) is the smallest algebraically closed field \(\overline{F}\) containing \(F\) as a subfield.

For example, \(\mathbb{R} = \mathbb{C}\). In fact, \(\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}\), i.e. it is a degree 2 field extension of \(\mathbb{R}\) obtained by adjoining \(i\), which is a root of \(1 + x^2\). This construction is very similar to the one we discussed in the previous section. We can write \(\mathbb{C}\) as a quotient ring \(\mathbb{C} = \mathbb{R}[x]/(1 + x^2)\). We remark that \(\overline{\mathbb{Q}}\) is contained in \(\mathbb{C}\), but in fact is smaller.

Next theorem describes the algebraic closure of \(\mathbb{F}_p\), and also of any finite field of characteristic \(p\). Similarly to Example 2.12, we have a chain of subfields of characteristic \(p\):
\[ \mathbb{F}_p \subset \mathbb{F}_{p^2} \subset \mathbb{F}_{p^3} \subset \cdots \subset \mathbb{F}_{p^{(n-1)!}} \subset \mathbb{F}_{p^n} \subset \cdots \]

Theorem 2.15.
\[ \overline{\mathbb{F}_p} = \bigcup_{n=1}^{\infty} \mathbb{F}_{p^n}. \]

Proof. Let \(K = \bigcup_{n=1}^{\infty} \mathbb{F}_{p^n}\). Clearly \(\mathbb{F}_p \subset K\). It is easy to see that \(K\) is a field. Indeed, for any \(\alpha, \beta \in K\) there exists \(k \geq 1\) such that \(\alpha, \beta \in \mathbb{F}_{p^k}\). Since the field axioms are satisfied for \(\alpha, \beta\) in \(\mathbb{F}_{p^k}\), they are also satisfied in \(K\).

To show \(K\) is algebraically closed consider any polynomial \(f \in K[x]\). Again, each of the coefficients of \(f\) lies in some finite field in the above union, so we can choose \(k \geq 1\) such that all coefficients of \(f\) lie in \(\mathbb{F}_{p^k}\), i.e. \(f \in \mathbb{F}_{p^k}[x]\). Now let \(\alpha\) be a root of \(f\). Then we obtain a finite extension \(\mathbb{F}_{p^k} \subset \mathbb{F}_{p^k}(\alpha)\) of some degree \(d\). There exists \(n\) such that \(kld\) divides \(n!\). Therefore \(\mathbb{F}_{p^k}(\alpha) \subset \mathbb{F}_{p^n} \subset K\). This shows that \(\alpha \in K\).

Finally, to show that \(K\) is the smallest algebraically closed field containing \(\mathbb{F}_p\), note that \(\mathbb{F}_p\) must contain \(\mathbb{F}_{p^n}\) for any \(n\) since \(\overline{\mathbb{F}_p}\) must contain roots of irreducible polynomials over \(\mathbb{F}_p\) of degree \(n!\). Therefore, \(\overline{\mathbb{F}_p}\) must contain and, hence, equal to \(K\). \(\square\)
2.1.3. Polynomial Rings. Now we will review what we know about polynomials in one variable and see what remains true for polynomials in several variables.

Let \( R \) be a commutative ring with 1, and \( R[x] \) the ring of univariate polynomials with coefficients in \( R \). When \( R = \mathbb{F} \) is a field the following facts about \( \mathbb{F}[x] \) hold:

1. \( \mathbb{F}[x] \) is a PID (principal ideal domain).
2. \( \mathbb{F}[x] \) is a Euclidean domain, there is a Euclidean Algorithm in \( \mathbb{F}[x] \).
3. \( \mathbb{F}[x] \) is a UFD (unique factorization domain).
4. (Euclid’s lemma) If \( p \in \mathbb{F}[x] \) is irreducible and \( p|fg \) then either \( p|f \) or \( p|g \).

Moreover when \( R = \mathbb{Z} \) we have

5. (Gauss’ lemma) If \( p \in \mathbb{Z}[x] \) factors in \( \mathbb{Q}[x] \) then it factors in \( \mathbb{Z}[x] \).

How do we define \( \mathbb{F}[x, y] \), the ring of polynomials in two variables? One way would be to say that \( \mathbb{F}[x, y] \) consists of finite linear combinations of monomials \( x^iy^j \) with coefficients in \( \mathbb{F} \) and \( i \geq 0, j \geq 0 \):

\[
f(x, y) = \sum_{i,j \geq 0} a_{i,j}x^iy^j, \quad a_{i,j} \in \mathbb{F}, \text{ all but finitely many } a_{i,j} \text{ are zero.}
\]

Another way is to set \( R = \mathbb{F}[x] \), which is a commutative ring with 1, and define \( \mathbb{F}[x, y] = R[y] = \mathbb{F}[x][y] \). In other words, \( \mathbb{F}[x, y] \) consists of polynomials in \( y \) with coefficients being polynomials in \( x \). You should show that the two definitions are equivalent.

Although we will mostly be dealing with bivariate polynomials we will give the general definition of the multivariate polynomial ring.

**Definition 2.16.** Let \( R \) be a commutative ring with 1. Define a \( n \)-variate polynomial \( f(x_1, \ldots, x_n) \) over \( R \) as a finite linear combinations of monomials \( x_1^{i_1} \cdots x_n^{i_n} \) with coefficients in \( R \):

\[
f(x_1, \ldots, x_n) = \sum_{i_1, \ldots, i_n \geq 0} a_{i_1, \ldots, i_n}x_1^{i_1} \cdots x_n^{i_n}, \quad a_{i_1, \ldots, i_n} \in R,
\]

where all but finitely many \( a_{i_1, \ldots, i_n} \) are zero. The set of all polynomials \( f(x_1, \ldots, x_n) \) forms the ring of polynomials \( R[x_1, \ldots, x_n] \) under usual operations of addition and multiplication. The (total) degree of a monomial \( x_1^{i_1} \cdots x_n^{i_n} \) is \( i_1 + \cdots + i_n \). The (total) degree \( \text{deg} \ f \) of a polynomial \( f \in R[x_1, \ldots, x_n] \) is the largest degree of monomials appearing in \( f \).

Equivalently, we can define \( R[x_1, \ldots, x_n] \) by induction on the number of variables: first define \( R[x] \), then define \( R[x_1, \ldots, x_n] \) as \( R[x_1][x_2, \ldots, x_n] \).

Here are some simple properties of degree, which you are invited to check yourself.

**Proposition 2.17.** For any \( f, g \) in \( R[x_1, \ldots, x_n] \) we have

1. \( \text{deg}(fg) = \text{deg} f + \text{deg} g \).
2. \( \text{deg}(f + g) \leq \max\{\text{deg} f, \text{deg} g\} \).

The notion of irreducibility is the same as for \( \mathbb{F}[x] \).

**Definition 2.18.** A non-constant polynomial \( f \in \mathbb{F}[x_1, \ldots, x_n] \) is called reducible over \( \mathbb{F} \) if \( f = gh \) for some non-constant polynomials \( g, h \in \mathbb{F}[x_1, \ldots, x_n] \). In this case we will also say that \( f \) factors in \( \mathbb{F}[x_1, \ldots, x_n] \). If \( f \) is not reducible over \( \mathbb{F} \) it is called irreducible over \( \mathbb{F} \).

**Example 2.19.**
(a) The polynomial \( x^2 - y^2 \in \mathbb{Q}[x,y] \) is reducible over \( \mathbb{Q} \) (and hence over \( \mathbb{R} \), and \( \mathbb{C} \)), since \( x^2 - y^2 = (x - y)(x + y) \), where \( x - y, x + y \in \mathbb{Q}[x,y] \).
(b) The polynomial \( x^2 + y^2 \) is reducible over \( \mathbb{C} \), since \( x^2 + y^2 = (x - iy)(x + iy) \). However, it is irreducible over \( \mathbb{R} \) (and hence over \( \mathbb{Q} \)). Indeed, suppose \( x^2 + y^2 = g(x,y)h(x,y) \) for non-constant \( g, h \in \mathbb{R}[x,y] \). Then by the property of degree both \( g \) and \( h \) are linear. Without loss of generality we may assume that the coefficient of \( y \) in each of them equals 1, so we have:

\[
x^2 + y^2 = (a_0 + a_1x + y)(b_0 + b_1x + y).
\]

Comparing the coefficients of \( x^2 \) and \( xy \) on both sides we get a system:

\[
1 = a_1b_1 \quad \text{and} \quad 0 = a_1 + b_1 \quad \text{which implies} \quad b_1^2 = -1.
\]

This is impossible for \( b_1 \in \mathbb{R} \) so no such non-constant \( g, h \in \mathbb{R}[x,y] \) exist.

(c) The polynomial \( x^2 + y^2 - 1 \) is irreducible over \( \mathbb{C} \) (and hence over \( \mathbb{R} \) and \( \mathbb{Q} \)) and we will show this later.

Let us return to the facts (1)–(5) that hold for univariate polynomials. This time the situation is a bit different.

1. \( \mathbb{F}[x,y] \) is not a PID.
2. \( \mathbb{F}[x,y] \) is not a Euclidean domain.
3. \( \mathbb{F}[x,y] \) is a UFD.
4. (Euclid’s lemma) If \( p \in \mathbb{F}[x,y] \) is irreducible and \( p|fg \) then either \( p|f \) or \( p|g \) in \( \mathbb{F}[x,y] \).
5. (Gauss’ lemma) A polynomial \( f \) is irreducible in \( \mathbb{F}[x,y] \) if and only if it is irreducible in \( \mathbb{F}(x)[y] \).

In the last statement \( \mathbb{F}(x) \) denotes the field of rational functions (quotients of polynomials) over \( \mathbb{F} \).

For example, the ideal \( I = \langle x, y \rangle \) generated by \( x \) and \( y \) in \( \mathbb{F}[x,y] \) is not principal. By definition, \( \langle x, y \rangle = \{ xh + yg \mid h, g \in \mathbb{F}[x,y] \} \). Suppose there exists \( f \in \mathbb{F}[x,y] \) such that \( I = \langle f \rangle \). Since \( I \) is proper \( f \) cannot be a constant. Then \( x = h_1f \) and \( y = h_2f \) for some \( h_1 \in \mathbb{F}[x,y] \). Comparing the degrees we see that \( \deg f = 1 \) and \( h_i \) are constants. But that means \( h_1^{-1}x = h_2^{-1}y \), a contradiction.

A general fact from abstract algebra says that Euclidean domains are PIDs, so \( \mathbb{F}[x,y] \) is not a Euclidean domain. Below we will prove the unique factorization property of \( \mathbb{F}[x,y] \) assuming Euclid’s lemma. The proof will be complete after we prove Gauss’s lemma and Euclid’s lemma in the next subsection.

**Theorem 2.20. (Unique Factorization) \( \mathbb{F}[x,y] \) is a UFD, i.e. every non-constant polynomial \( f \in \mathbb{F}[x,y] \) can be written as a product \( f = f_1 \cdots f_s \) where each \( f_i \) is irreducible over \( \mathbb{F} \). This product is unique up to ordering the factors and multiplying by constants.**

For example, \( f = f_1f_2 = f_2f_1 = (cf_1)(\frac{1}{c}f_2) \) is considered to be the same factorization up to ordering the factors and multiplying by constants.

**Proof.** The existence of the factorization is easy to see by induction on the degree of \( f \). Indeed, if \( f \) is irreducible, we are done. Otherwise \( f \) factors \( f = gh \) for some non-constant \( g, h \in \mathbb{F}[x,y] \) of smaller degree than \( \deg f \). By the inductive hypothesis both \( g \) and \( h \) factor into irreducible factors and, hence, so does \( f \).

For uniqueness, suppose

\[
f = f_1 \cdots f_s = g_1 \cdots g_t,
\]
for irreducible \( f_i \) and \( g_i \) in \( \mathbb{F}[x,y] \). By Euclid's lemma \( f_1 \) must divide one of the \( g_i \), up to reordering we may assume that \( f_1 \parallel g_1 \). Since \( g_1 \) is also irreducible we obtain

\[
f_1 = c_1 g_1
\]

for some constant \( c_1 \). Now since \( \mathbb{F}[x,y] \) has no zero divisors we can cancel \( f_1 \) and obtain

\[
f_2 \cdots f_s = c_1^{-1} g_2 \cdots g_t.
\]

Continuing in this way we see that \( t = s \) and \( f_i = g_i \) up to ordering and multiplying by constants.

\[\square\]

2.1.4. Gauss's Lemma and Euclid's Lemma. We will start with a definition.

**Definition 2.21.** A polynomial \( f \) in \( \mathbb{F}[x][y] \) is called primitive if its coefficients \( a_i(x) \) are relatively prime as elements of \( \mathbb{F}[x] \), i.e. if \( f(x,y) = \sum_{i=1}^{k} a_i(x)y^i \) then \( \gcd(a_i(x) \mid 1 \leq i \leq k) = 1 \).

**Lemma 2.22.** If \( f, g \) in \( \mathbb{F}[x][y] \) are primitive then so is \( fg \).

**Proof.** Assume it is not, i.e. \( fg = c(x)h \) for some non-constant \( c(x) \in \mathbb{F}[x] \) and \( h \in \mathbb{F}[x][y] \). Let \( p(x) \) be an irreducible factor of \( c(x) \). The identity \( fg = c(x)h \) in \( (\mathbb{F}[x]/\langle p(x) \rangle)[y] \) becomes \( \bar{f}\bar{g} = 0 \). This implies that either \( \bar{f} = 0 \) or \( \bar{g} = 0 \) in \( (\mathbb{F}[x]/\langle p(x) \rangle)[y] \) (remember that this is a polynomial ring over a field, hence it has no zero divisors). But this means that \( p(x) \) divides every coefficient of either \( f \) or \( g \), i.e. either \( f \) or \( g \) is not primitive.

Recall that \( \mathbb{F}(x) \) denotes the field of rational functions in \( x \) over \( \mathbb{F} \), i.e.

\[\mathbb{F}(x) = \left\{ \frac{f(x)}{g(x)} \mid f, g \in \mathbb{F}[x], g \neq 0 \right\} .\]

**Theorem 2.23.** (Gauss's Lemma) A polynomial \( f \) is irreducible in \( \mathbb{F}[x,y] \) if and only if it is irreducible in \( \mathbb{F}(x)[y] \).

**Proof.** \((\leftarrow)\) If \( f \) has a non-trivial factorization in \( \mathbb{F}[x,y] \) then this factorization makes sense in \( \mathbb{F}(x)[y] \) as well.

\((\rightarrow)\) Suppose \( f \) is irreducible in \( \mathbb{F}[x,y] \), but has a non-trivial factorization

\[f(x,y) = g(x,y)h(x,y), \quad \text{for some } g, h \in \mathbb{F}(x)[y]\]

First, note that \( f \) is primitive as an element of \( \mathbb{F}(x)[y] \), since \( f \) is irreducible in \( \mathbb{F}[x,y] \). We can clear the denominators in \( g \) and \( h \), i.e. find \( a, b \in \mathbb{F}[x] \) such that \( a(x)g(x,y) \) and \( b(x)h(x,y) \) lie in \( \mathbb{F}[x][y] \). Let us factor out the gcd's of their coefficients to make them primitive:

\[a(x)g(x,y) = c(x)g_1(x,y), \quad b(x)h(x,y) = d(x)h_1(x,y),\]

where \( g_1 \) and \( h_1 \) lie in \( \mathbb{F}[x][y] \) and are primitive. From (2.1) we obtain

\[a(x)b(x)f(x,y) = c(x)d(x)g_1(x,y)h_1(x,y).\]

Now \( f(x,y) \) is primitive and by Lemma 2.22 the product \( g_1(x,y)h_1(x,y) \) is also primitive, hence, \( a(x)b(x) = c(x)d(x) \) in \( \mathbb{F}[x] \). Therefore \( f(x,y) = g_1(x,y)h_1(x,y) \) is a non-trivial factorization in \( \mathbb{F}[x][y] \), which contradicts the irreducibility of \( f \).

\[\square\]

Now we can prove Euclid's lemma.

**Theorem 2.24.** (Euclid's Lemma) Let \( f \in \mathbb{F}[x,y] \) be irreducible. Then \( f|gh \) implies \( f|g \) or \( f|h \) in \( \mathbb{F}[x,y] \).
PROOF. Consider \( f, g, h \) as elements of \( \mathbb{F}(x)[y] \). By Gauss’s lemma \( f \) is irreducible in \( \mathbb{F}(x)[y] \). Now \( \mathbb{F}(x)[y] \) is the ring of univariate polynomials over a field, so by the usual Euclid’s lemma \( f|gh \) implies \( f|g \) or \( f|h \) in \( \mathbb{F}(x)[y] \). We will assume the former, so

\[
g(x, y) = f(x, y)q(x, y) \quad \text{for some } q \in \mathbb{F}(x).[y].
\]

(2.2)

Let’s clear the denominators: there exists \( c(x) \in \mathbb{F}[x] \) such that \( c(x)q(x, y) \in \mathbb{F}[x, y] \). As before, factor out the gcd of its coefficients to make it primitive: \( c(x)q(x, y) = d(x)q_1(x, y) \) for a primitive \( q_1 \in \mathbb{F}[x][y] \). From (2.2) we get

\[
c(x)g(x, y) = f(x, y)d(x)q_1(x, y).
\]

Both \( f(x, y) \) and \( q_1(x, y) \) are primitive, and so is their product, by Lemma 2.22. Therefore, \( d(x) \) is the gcd of the coefficients on the right hand side, which implies that \( c(x)d(x) \) in \( \mathbb{F}[x] \). Now

\[
g(x, y) = f(x, y)\left(\frac{d(x)}{c(x)}q_1(x, y)\right)
\]

is a factorization in \( \mathbb{F}[x][y] \) which shows that \( f|g \) in \( \mathbb{F}[x, y] \). \( \square \)

Remark 2.25. What we said about bivariate polynomials in (1)–(5) above is also true for polynomials in any number of variables. In fact, one can adapt all our proofs to the general case (e.g. use induction on the number of variables).

2.1.5. Eisenstein Criterion. You may have seen the Eisenstein criterion for irreducibility of polynomials in \( \mathbb{Z}[x] \):

**Theorem 2.26.** Let \( f = a_0 + a_1x + \cdots + a_nx^n \in \mathbb{Z}[x] \). If there exists a prime \( p \) such that

(i) \( p|a_i \) for \( 0 \leq i \leq n-1 \),

(ii) \( p \nmid a_n \),

(iii) \( p^2 \nmid a_0 \)

then \( f \) is irreducible over \( \mathbb{Q} \).

We will prove a similar criterion for polynomials in \( \mathbb{F}[x, y] \). First, a definition.

**Definition 2.27.** A polynomial \( f \in \mathbb{F}[x, y] \) is called absolutely irreducible if \( f \) is irreducible in \( \mathbb{K}[x, y] \) for any finite extension \( \mathbb{F} \subset \mathbb{K} \).

**Example 2.28.**

(a) \( x^2 + y^2 \in \mathbb{R}[x, y] \) is not absolutely irreducible, since it is reducible over \( \mathbb{C} \), which is a finite extension of \( \mathbb{R} \).

(b) Any linear polynomial in \( \mathbb{F}[x, y] \) is absolutely irreducible.

(c) \( x^2 + y^2 - 1 \in \mathbb{R}[x, y] \) is absolutely irreducible, which we will see in a moment.

**Proposition 2.29.** Let \( f \in \mathbb{F}[x, y] \) be non-constant. Then there exists a finite extension \( \mathbb{F} \subset \mathbb{K} \) such that \( f \) factors into a product of absolutely irreducible polynomials in \( \mathbb{K}[x, y] \).

**Proof.** Induction on \( n = \deg f \). The base case is when \( f \) is linear, and hence is already absolutely irreducible. Suppose \( n > 1 \). If \( f \) is absolutely irreducible we are done. Otherwise, there exists a finite extension \( \mathbb{F} \subset \mathbb{L} \) such that \( f = f_1f_2 \) for some non-constant \( f_1, f_2 \) in \( \mathbb{L}[x, y] \) of degree smaller than \( n \). By the inductive
hypothesis there exist finite extensions $L \subset K_1$ and $L \subset K_2$ such that $f_i$ is a product of absolutely irreducible factors in $K_i[x, y]$. Let $K = K_1 + K_2$, which is a finite extension of $L$, and hence of $F$. Then $f$ factors into a product of absolutely irreducible polynomials in $K[x, y]$.

We are ready for the Eisenstein Criterion.

**Theorem 2.30. (The Eisenstein Criterion)** Let $f = a_0(x) + a_1(x)y + \cdots + a_n(x)y^n$ be a primitive non-constant polynomial in $F[x, y]$. Suppose there exists $\alpha$ in some finite extension $K$ of $F$ such that

(i) $\alpha$ is a root of $a_i(x)$ for $0 \leq i \leq n - 1$,
(ii) $\alpha$ is not a root of $a_n(x)$, 
(iii) $\alpha$ is not a multiple root $a_0(x)$.

Then $f$ is absolutely irreducible.

**Proof.** Suppose not. Then there exists a finite extension $F \subset L$ such that $f = gh$ for some non-constant $g, h$ in $L[x, y]$. We may assume that $a$ lies in $L$ (otherwise replace $L$ with $L + K$). We have

$$g(x, y) = \sum_{i=0}^{k} b_i(x)y^i, \quad h(x, y) = \sum_{i=0}^{l} c_i(x)y^i, \quad \text{where } k + l = n.$$  

Then

$$f(x, y) = g(x, y)h(x, y) = b_0(x)c_0(x) + \cdots + b_k(x)c_l(x)y^n.$$  

This implies that $a_0(x) = b_0(x)c_0(x)$, and so either $b_0(\alpha) = 0$ or $c_0(\alpha) = 0$ by (i). Without loss of generality we may assume that $b_0(\alpha) = 0$. Then $c_0(\alpha) \neq 0$, otherwise $\alpha$ would be a multiple root of $a_0(x)$, which we assumed is not by (iii).

Also $a_n(x) = b_k(x)c_l(x)$, and so $b_k(\alpha) \neq 0$ and $c_l(\alpha) \neq 0$ by (ii). Let $s$ be the smallest $s \leq k$ such that $b_s(\alpha) \neq 0$. If $s < n$ then

$$a_s(x) = \sum_{i+j=s} b_i(x)c_j(x) = b_0(x)c_s(x) + b_1(x)c_{s-1}(x) + \cdots + b_s(x)c_0(x).$$  

Plugging in $x = \alpha$ we obtain $0 = b_s(\alpha)c_0(\alpha)$, which is a contradiction since neither $b_s(\alpha)$ nor $c_0(\alpha)$ is zero. Therefore $s = n = k$, which means that $l = 0$ and so $h(x, y) = c_0(x)$ and $f(x, y) = c_0(x)g(x, y)$. But we assumed that $f$ was primitive, hence $h(x, y) = c_0(x) = c_0$, a constant. This shows that $f$ must be absolutely irreducible. \qed

**Example 2.31.**

(a) $f(x, y) = x^2 + y^2 - 1 \in \mathbb{R}[x, y]$ is absolutely irreducible. Indeed, $f(x, y) = (x^2 - 1) + y^2$, i.e. $a_0(x) = x^2 - 1$, $a_1(x) = 0$, and $a_2(x) = 1$. Choose $a = 1$, then (i) $a_0(1) = 0$ and $a_1(1) = 0$, (ii) $a_2(1) \neq 0$, and (iii) $a_0'(1) = 2 \neq 0$. By the Eisenstein criterion $x^2 + y^2 - 1$ is absolutely irreducible.

(b) Consider $y^n - f(x) \in F[x, y]$ where $f(x)$ is a polynomial which has a simple (i.e. non-multiple) root in some finite extension of $F$. Then $y^n - f(x)$ is absolutely irreducible. For example, $y^n - x$ is absolutely irreducible. Note a striking distinction between univariate and multivariate case. When $F$ is algebraically closed, the only irreducible univariate polynomials are linear, whereas there are absolutely irreducible multivariate polynomials of arbitrarily large degree.
2.2. Affine and Projective Curves

2.2.1. Affine Curves. Let \( \mathbb{F} \) be a field. We call the Cartesian power
\[
\mathbb{F}^n = \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{F}\}
\]
the \textit{affine space over} \( \mathbb{F} \) and denote by \( \mathbb{A}_2^g \) or, simply, by \( \mathbb{A}^n \). In particular, \( \mathbb{A}^2 \) is called the affine plane and \( \mathbb{A}^1 \) is the affine line. This may look redundant, but it will be handy later when we talk about the affine and the projective plane.

**Definition 2.32.** A \textit{plane affine curve} \( C \) is the set
\[
C = \{(x, y) \in \mathbb{A}^2 \mid f(x, y) = 0\}
\]
for some non-constant polynomial \( f \in \mathbb{F}[x, y] \). The \textit{degree of} \( C \) is the degree of the polynomial \( f \). It is custom to call curves of degree two \textit{conics} and curves of degree three \textit{cubics}.

Let us write \( f \) as a product of distinct (absolutely) irreducible factors
\[
f = f_1^{k_1} \cdots f_s^{k_s},
\]
where \( f_i \neq cf_j \) for any \( i \neq j, c \in \mathbb{F} \).

Then \( C \) is a union of curves
\[
C_i = \{(x, y) \in \mathbb{A}^2 \mid f_i(x, y) = 0\},
\]
which are called the \textit{(absolutely) irreducible components of} \( C \). A curve with only one (absolutely) irreducible component is called \textit{(absolutely) irreducible curve}.

**Example 2.33.** Let \( \mathbb{F} = \mathbb{R} \), the real numbers.
(a) \( f(x, y) = x^2 - y^2 = (x - y)(x + y) \). The curve \( C \) is the union of two lines \( y = x \) and \( y = -x \), which are the absolutely irreducible components of \( C \).
(b) \( f(x, y) = x^2 + y^2 - 1 \). The curve \( C \) is the unit circle. It has only one absolutely irreducible component according to part (a) of Example 2.31.
(c) \( f(x, y) = a_0(x) \). The irreducible components of \( C \) are vertical lines \( x = \alpha \), for every real root \( \alpha \) of \( a_0(x) \). The absolutely irreducible components are the vertical lines \( x = \alpha \) for every complex root \( \alpha \) of \( a_0(x) \).

We would like to have a one-to-one correspondence between irreducible curves \( C \) and their defining polynomials \( f \in \mathbb{F}[x, y] \), up to a constant multiple. Here is a problem, though: the “curve” in \( \mathbb{A}^2_\mathbb{R} \) defined by \( f(x, y) = x^2 + y^2 \) consists of just the origin \( C = \{(0, 0)\} \). The same curve can be defined by many other polynomials, e.g. \((2x + y)^2 + (x - 2y)^2\) or \((y - x^2)^2 + x^2\), etc. This difficulty can be resolved if we consider curves over algebraically closed fields. From now on we will assume that \( \mathbb{K} \) denotes and algebraically closed field, whereas \( \mathbb{F} \) will denote an arbitrary field.

**Proposition 2.34.** If \( \mathbb{K} \) is algebraically closed then any curve \( C \) defined by \( f \in \mathbb{K}[x, y] \) has infinitely many points.

**Proposition 2.35.** If \( f, g \in \mathbb{F}[x, y] \) such that \( f \) is irreducible over \( \mathbb{F} \) and \( f \) does not divide \( g \). Then the set of common their zeroes
\[
\{(x, y) \in \mathbb{A}^2 \mid f(x, y) = 0, g(x, y) = 0\}
\]
is finite.
The following theorem establishes the one-to-one correspondence we discussed above.

**Theorem 2.36.** Let \( C \subset \mathbb{A}^2_\mathbb{K} \) be an irreducible curve defined by an irreducible polynomial \( f \in \mathbb{K}[x,y] \). Then \( C \) determines \( f \) uniquely up to a constant multiple.

**Proof.** According to Proposition 2.34, \( C \) has infinite number of points. Suppose \( g \in \mathbb{K}[x,y] \) is another irreducible polynomial defining \( C \). Then \( C \) is the set of common zeroes of \( f \) and \( g \) and by Proposition 2.35 \( f \) must divide \( g \). Since both \( f, g \) are irreducible we must have \( g = cf \) for some \( c \in \mathbb{K} \).

The proof of Proposition 2.34 relies on the fact that an algebraically closed field \( \mathbb{K} \) must be infinite. This is not hard to see: if \( \mathbb{K} \) was finite, \( \mathbb{K} = \{0, 1, \alpha_3, \ldots, \alpha_n\} \), we could write down a polynomial, say, \( f(x) = x(x-1)(x-\alpha_1)\cdots(x-\alpha_n)+1 \) which has no roots in \( \mathbb{K} \).

**Proof of Proposition 2.34.** Suppose \( C \) is defined by \( f(x,y) = a_0(x) + \cdots + a_n(x)y^n \in \mathbb{K}[x,y], \) for \( n \geq 1 \).

For any \( \alpha \in \mathbb{K} \) the polynomial \( f(\alpha, y) \) lies in \( \mathbb{K}[y] \) and hence must have \( n \) roots, counting multiplicities. Since \( \mathbb{K} \) is infinite we obtain infinitely many points \((\alpha, \beta)\) on \( C \), where \( \alpha \) is arbitrary and \( \beta \) is root of \( f(\alpha,y) \). The case \( n = 0 \) is left for you.

**Proof of Proposition 2.35.** By Gauss’s lemma \( f \) is irreducible in \( \mathbb{F}(x)[y] \). Also \( f \) does not divide \( g \) in \( \mathbb{F}(x)[y] \) (we saw in the proof of Euclid’s lemma that if \( f|g \) in \( \mathbb{F}(x)[y] \) then \( f|g \) in \( \mathbb{F}[x][y] \)). Therefore, \( \text{gcd}(f,g) = 1 \), i.e. there exist \( u, v \in \mathbb{F}(x,y) \) such that \( uf + vg = 1 \). After clearing the denominators in \( u, v \) we get \( u_1 f + v_1 g = c(x) \) for some \( c(x) \in \mathbb{F}[x] \) and \( u_1, v_1 \in \mathbb{F}[x,y] \). If \((\alpha, \beta)\) is a common zero of \( f,g \) then \( c(\alpha) = 0 \). Therefore there could be only finitely many such \( \alpha \). Furthermore, \( \beta \) must be a root of \( f(\alpha,y) \), so there are only finitely many such \( \beta \) as well.

It is an interesting question how many common zeroes \( f \) and \( g \) can have. We will answer this question in the case of an algebraically closed field. We will see that the number of common zeroes is at most the product of the degrees of \( f, g \). In fact, Bézout’s theorem (see Theorem 2.65) says that it is always the product of the degrees if we count the common zeroes not in the affine plane, but in a compact space, called the projective plane. This is the subject of the next subsection.

**2.2.2. Projective Plane.** We will start with the following question: How many times does a line \( L \) intersect a plane affine curve \( C \) of degree \( n \)? We will see that the answer is at most \( n \) unless \( L \) is an irreducible component of \( C \).

**Theorem 2.37.** Let \( C \) be a plane affine curve of degree \( n \). Then any line \( L \) intersects \( C \) in at most \( n \) points, unless \( L \subset C \).

**Proof.** Let \( f, l \in \mathbb{F}[x,y] \) be a degree \( n \) polynomial and a linear polynomial defining \( C \) and \( L \), respectively. If \( l \) divides \( f \) then \( L \) is an irreducible component of \( C \), otherwise \( C \cap L \) is finite. Let \( l(x,y) = ay + bx + c \) and assume \( a \neq 0 \). Then the \( x \)-coordinates of the points of \( C \cap L \) are the roots of \( f(x, -\frac{b}{a} x - \frac{c}{a}) \). This is a polynomial of degree at most \( n \), and hence, has at most \( n \) roots. Indeed, every monomial \( x^k y^m \) in \( f \) of degree \( k + m \leq n \) produces a polynomial \( x^k \left( -\frac{b}{a} x - \frac{c}{a} \right)^m \) of degree at most \( n \). If \( \alpha_1, \ldots, \alpha_s, s \leq n \), are the roots of \( f(x, -\frac{b}{a} x - \frac{c}{a}) \), then
\[ C \cap L = \{(\alpha_i, -\frac{b}{a}\alpha_i - \frac{c}{a}) \mid 1 \leq i \leq s\} \]. The case \( a = 0 \) is left for you as an exercise.

\[ \text{QUESTION. Why can } |C \cap L| \text{ be strictly smaller than } \deg C? \]

\[ \text{ANSWER. In the above proof we saw two reasons for this. First, a polynomial of degree } n \text{ may have fewer than } n \text{ roots if the field is not algebraically closed. Second, it may happen that the degree of } f(x, -\frac{b}{a}x - \frac{c}{a}) \text{ is strictly smaller than } \deg f \text{ as there could be cancelation of highest degree terms.} \]

We will illustrate both reasons using the following example.

**Example 2.38.** Let \( C \) be the parabola defined by \( f(x, y) = x^2 - y \in \mathbb{R}[x, y] \). Consider any line \( L \) with equation \( ay + bx + c = 0 \) and assume \( a \neq 0 \). The \( x \)-coordinates of the intersection points are given by \( ax^2 + bx + c = 0 \). This polynomial has two real roots if \( b^2 - 4ac > 0 \), one multiple real root if \( b^2 - 4ac = 0 \), and two complex roots if \( b^2 - 4ac < 0 \). Thus, if we consider \( C \) over complex numbers then the number of intersections with non-vertical line is always 2, which is the degree of \( C \). Note that this includes the case when \( L \) is tangent to \( C \). In this case the intersection point, is a double intersection point (intersection with multiplicity 2).

If \( a = 0 \) the line \( L \) is vertical and has equation \( bx + c = 0 \). It intersects \( C \) only once: the intersection point is \( (\frac{c}{b}, \frac{c^2}{b^2}) \). Imagine now \( a \) is very small, then the line \( L \) is almost vertical, in fact one of the intersection points is very close to \( (\frac{c}{b}, \frac{c^2}{b^2}) \) and the other one is “very far” on the parabola, i.e. its \( y \)-coordinate is very large. So if we equip \( C \) with one extra point “at infinity” then any vertical line with intersect \( C \) at two points, one of which is this extra point. Notice that this way we obtain a compact curve, the limit of a sequence of points on \( C \) with the \( y \)-coordinate approaching infinity is the infinite point we added to \( C \).

A similar situation occurs when we consider the intersection of two lines. The number of intersection points is always one, except when they are parallel. If we change one of the lines slightly the intersection point will appear “very far” in the direction of these almost parallel lines. If we add a “point at infinity” in the direction of two parallel lies then we can still say that the two lines intersect at this extra point.

\[ \text{Figure 2.1. Intersection with a line: (a) two real points, (b) two complex points, (c) one real, one infinite points} \]
To make the above arguments more precise and formal we will define the projective plane and will be considering curves in the projective plane. First, let us define the projective line. In the $xy$-plane consider the set of all lines through $(0,0)$. They intersect the line $y = 1$ at a point $(u,1)$. This defines a one-to-one correspondence between points on the line $y = 1$ and all non-horizontal lines passing through $(0,0)$.

![Figure 2.2. Constructing the projective line](image)

**Definition 2.39.** The *projective line* $\mathbb{P}^1_F$ over $F$ is the set of all lines in $\mathbb{A}^2_F$ passing through the origin $(0,0)$. We will also write $\mathbb{P}^1$ without specifying the field.

Using the correspondence described above we can identify points of $\mathbb{P}^1$ with points of $\mathbb{A}^1 \cup \{\infty\}$, where the non-horizontal lines correspond to the points of $\mathbb{A}^1$ and the horizontal line $y = 0$ corresponds to one extra point which we denote by $\infty$. Note that $\mathbb{A}^1$ is naturally embedded into $\mathbb{P}^1$; the embedding is given by the map

$$\mathbb{A}^1 \hookrightarrow \mathbb{P}^1, \quad u \mapsto \text{line connecting } (0,0) \text{ and } (u,1).$$

It is convenient to have a coordinate system on $\mathbb{P}^1$, just like we have on $\mathbb{A}^1$. The idea is that every point $(x,y) \in \mathbb{A}^2 \setminus \{(0,0)\}$ defines a unique point in $\mathbb{P}^1$, namely the line connecting $(0,0)$ and $(x,y)$. However two points $(x_1,y_1)$ and $(x_2,y_2)$ in $\mathbb{A}^2 \setminus \{(0,0)\}$ will define the same point in $\mathbb{P}^1$ if and only if $(x_2,y_2) = \lambda (x_1,y_1)$ for some non-zero constant $\lambda \in F$. We obtain an equivalence relation on $\mathbb{A}^2 \setminus \{(0,0)\}$:

$$(x_1,y_1) \sim (x_2,y_2) \text{ if and only if } (x_2,y_2) = \lambda (x_1,y_1) \text{ for some } \lambda \in F^*.$$ 

This gives us an equivalent definition of the projective line:

$$\mathbb{P}^1 = \frac{\mathbb{A}^2 \setminus \{(0,0)\}}{\sim} = \{ \text{classes of pairs } (x,y) \in \mathbb{A}^2 \setminus \{(0,0)\} \text{ under } \sim \}.$$ 

We denote by $(x:y)$ the equivalence class of $(x,y)$ and call it the *homogeneous coordinate* of the corresponding point in $\mathbb{P}^1$. For example, $(1:2) = (1/2:1) = (10:20)$ correspond to the same point in $\mathbb{P}^1$.

Now we can generalize our definition.

**Definition 2.40.** The *projective $n$-space* $\mathbb{P}^n_F$ over $F$ is the set of all lines in $\mathbb{A}^{n+1}_F$ passing through the origin 0. We will also write $\mathbb{P}^n$ without specifying the field.
As before we have the embedding
\[ A^n \hookrightarrow \mathbb{P}^n, \quad (u_1, \ldots, u_n) \mapsto \text{line connecting } \mathbf{0} \text{ and } (u_1, \ldots, u_n, 1). \]

Similarly, we define the homogeneous coordinates on \( \mathbb{P}^n \). First, \((x_0, \ldots, x_n) \sim (y_0, \ldots, y_n)\) if and only if \((x_0, \ldots, x_n) = \lambda(y_0, \ldots, y_n)\) for some \(\lambda \in \mathbb{F}^*\) is an equivalence relation on \(A^{n+1} \setminus \{0\}\). We obtain
\[ \mathbb{P}^n = \frac{A^{n+1} \setminus \{0\}}{\sim} = \left\{ \text{classes of } (x_0, \ldots, x_n) \in A^{n+1} \setminus \{0\} \text{ under } \sim \right\}. \]

We denote by \((x_0 : \cdots : x_n)\) the equivalence class of \((x_0, \ldots, x_n)\) under \(\sim\), and call it the *homogeneous coordinate* of the corresponding point in \(\mathbb{P}^n\).

In particular, we have the *projective line*:
\[ \mathbb{P}^2 = \{ \text{lines in } \mathbb{A}^3 \text{ passing through } (0, 0, 0) \}. \]

Two points \((x_1, y_1, z_1) \sim (x_2, y_2, z_2)\) if and only if \((x_2, y_2, z_2) = \lambda(x_1, y_1, z_1)\) for some \(\lambda \in \mathbb{F}^*\), in which case they define the same line in \(\mathbb{A}^3\) through \((0, 0, 0)\). The expression \((x : y : z)\) denotes the equivalence class of \((x, y, z)\) under the equivalence relation \(\sim\).

![Figure 2.3. Constructing the projective plane.](image-url)
we see that this is a projective line. Therefore
\[ \mathbb{P}^2 = U_z \cup L_z \cong \mathbb{A}^2 \cup \mathbb{P}^1. \]

This shows that \( \mathbb{P}^2 \) is a way to compactify \( \mathbb{A}^2 \) by adding a projective line \( L_z \) “at infinity”. We can also see it from our geometric construction in Figure 2.3. There is a one-to-one correspondence between the non-horizontal lines in \( \mathbb{A}^3 \) though the origin and the points of the plane \( z = 1 \) (which we identify with \( \mathbb{A}^2 \)). Every horizontal line through the origin in the \( z = 0 \) plane is a point in \( L_z \) which we identify with \( \mathbb{P}^1 \).

Here is another way to interpret (2.3). The points of \( \mathbb{P}^2 \) are the points in \( \mathbb{A}^2 \) (the points with coordinates \((x : y : 1)\)) together with an “infinite point” for every direction determined by a pair \((x, y)\) defined up to a non-zero scalar multiple (the points with coordinates \((x : y : 0)\)).

We can identify two more coordinate lines in \( \mathbb{P}^2 \): the line \( L_x \) is given by \( x = 0 \) and the line \( L_y \) is given by \( y = 0 \). Note that the three coordinate lines intersect at three special points with coordinates \((0 : 0 : 1)\), \((1 : 0 : 0)\), and \((0 : 1 : 0)\). We draw this schematically in Figure 2.4.

Now we will look at our Example 2.38 from the viewpoint of the new definition.

**Example 2.41.** First, consider two parallel lines in \( \mathbb{A}^2 \):
\[
L_1 = \{(u, v) \in \mathbb{A}^2 \mid v - 2u = 0\}, \quad L_2 = \{(u, v) \in \mathbb{A}^2 \mid v - 2u - 2 = 0\}.
\]

Where \((u, v)\) are the affine coordinates. Recall that in \( U_z \) we have \( u = \frac{x}{z}, v = \frac{y}{z} \), so the two equations become \( \frac{y}{z} - 2\frac{x}{z} = 0 \) and \( \frac{y}{z} - 2\frac{x}{z} - 2 = 0 \). If we clear the denominators we obtain \( y - 2x = 0 \) and \( y - 2x - 2z = 0 \). Notice that now they make sense for \( z = 0 \) as well. Thus they define two lines in the projective space:
\[
\bar{L}_1 = \{(x : y : z) \in \mathbb{P}^2 \mid y - 2x = 0\}, \quad \bar{L}_2 = \{(x : y : z) \in \mathbb{P}^2 \mid y - 2x - 2z = 0\},
\]
which coincide with \( L_1 \) and \( L_2 \) on \( U_z \). However each of them has one extra point \((1 : 2 : 0)\) outside of \( U_z \). Therefore \( \bar{L}_1 \) and \( \bar{L}_2 \) intersect in \( \mathbb{P}^2 \) at one point \((1 : 2 : 0)\).
Example 2.42. Let us see how a parabola \( C \) given by \( v - u^2 = 0 \) and a vertical line \( L \) given by \( u = \alpha \) intersect when we view them in the projective plane. As before, put \( u = \frac{x}{z}, \ v = \frac{y}{z} \) and rewrite the equations in the homogeneous coordinates.

\[
\bar{C} = \{ (x : y : z) \in \mathbb{P}^2 \mid y - x^2 = 0 \}, \quad \bar{L} = \{ (x : y : z) \in \mathbb{P}^2 \mid x - \alpha z = 0 \}.
\]

To see the intersection points substitute \( x = \alpha z \) into the equation of the parabola. Factoring out \( z \) we obtain a system

\[
\begin{cases}
x = \alpha z \\
z(y - \alpha^2 z) = 0
\end{cases}
\]

If \( z \neq 0 \) (i.e. in the affine part \( U_z \)) we get one solution \( (\alpha z : \alpha^2 z : z) = (\alpha : \alpha^2 : 1) \).
If \( z = 0 \) (i.e. on the infinite line \( L_z \)) we get one more solution \( (0 : y : 0) = (0 : 1 : 0) \).

Note that \( y \neq 0 \) since \( (0 : 0 : 0) \) does not represent any point in \( \mathbb{P}^2 \).
Summarizing, the parabola \( \bar{C} \) and the line \( \bar{L} \) intersect in \( \mathbb{P}^2 \) at two points \( \{ (\alpha : \alpha^2 : 1), (0 : 1 : 0) \} \).

2.2.3. Projective Curves. We are ready to define curves in the projective plane, which we call projective curves. The following definitions formalize the way we were passing from the \((u \mapsto v)\)-coordinates to the homogeneous coordinates in Example 2.41 and Example 2.42.

Definition 2.43. A polynomial \( F(x, y, z) \) in \( \mathbb{F}[x, y, z] \) is called \textit{homogeneous of degree} \( n \) if every monomial appearing in \( F \) has (total) degree \( n \):

\[
F(x, y, z) = \sum_{i+j \leq n} a_{ij} x^i y^j z^{n-i-j}.
\]

Definition 2.44. A \textit{plane projective curve} \( C \) is the set

\[
C = \{ (x : y : z) \in \mathbb{P}^2 \mid F(x, y, z) = 0 \}
\]

for some non-constant homogeneous polynomial \( F \in \mathbb{F}[x, y, z] \). The \textit{degree of} \( C \) is the degree of the polynomial \( F \). A curve \( C \) is called \textit{irreducible} if \( F \) is irreducible in \( \mathbb{F}[x, y, z] \) (see Definition 2.18).

Just as for affine curve, every projective curve \( C \) is a union of finitely many irreducible curves, called the \textit{irreducible components of} \( C \).

Let \( f \in \mathbb{F}[u, v] \) be a polynomial of degree \( n \). The following transformation

\[
f(u, v) \leadsto F(x, y, z) = z^n f \left( \frac{x}{z}, \frac{y}{z} \right)
\]

is called the \textit{homogenization} of \( f \). Conversely, the transformation

\[
F(x, y, z) \leadsto F(x, 1) = f(x, y)
\]

is called the \textit{dehomogenization} of \( F \). These transformations allow us to

- Start with an affine curve and construct the corresponding projective curve.
- See the affine part of the projective curve defined by \( F(x, y, z) = 0 \).

Definition 2.45. Let \( C \subset \mathbb{P}^2 \) be a projective curve. The set \( C_A = C \cap U_z \) is called the \textit{affine part} of \( C \). It is defined by the polynomial \( f(x, y) = F(x, y, 1) \).
2.3. Tangent Lines and Singular Points

We begin with an example of a projective curve in \( \mathbb{P}^2 \), intersecting three different lines.

**Example 2.46.** We will find the intersection points of a conic \( C \) with equation \( x^2 + 4y^2 - z^2 = 0 \) and

(a) \( L_1 \) given by \( x - y = 0 \),
(b) \( L_2 \) given by \( x - 2z = 0 \),
(c) \( L_3 \) given by \( x - z = 0 \)

For \( C \cap L_1 \) substituting \( x = y \) into the equation of \( C \) we get \( x^2 + 4x^2 - z^2 = 0 \) which factors as \((\sqrt{5}x - z)(\sqrt{5} + z) = 0\). This produces two points \((1 : 1 : \sqrt{5})\) and \((1 : 1 : -\sqrt{5})\) in \( \mathbb{P}^2 \).

Similarly, for \( C \cap L_2 \) we have \( 4z^2 + 4y^2 - z^2 = 0 \) which factors over complex numbers as \((2y - i\sqrt{3}z)(2y + i\sqrt{3}z) = 0\). We obtain two points \((4 : \sqrt{3}i : 2)\) and \((4 : -\sqrt{3}i : 2)\) in \( \mathbb{P}^2 \).

Finally, for \( C \cap L_3 \) we have \( z^2 + 4y^2 - z^2 = 0 \) which simplifies to \( y^2 = 0 \). This time we have a multiple solution \((1 : 0 : 1)\) with multiplicity two. Geometrically this means that \( L_3 \) is tangent to \( C \) at \((1 : 0 : 1)\).

In Figure 2.5 we depicted the affine real part of \( C \) and the lines \( L_i \), i.e. their intersection with \( U_z \) where the points have real coordinates. This can be easily obtained by dehomogenizing the equations of the curve and the lines.

![Figure 2.5. Intersection of a conic and three lines.](image)

**2.3.1. Tangent Lines.** By definition a line \( L \) is tangent to a curve \( C \) at a point \( p_0 \) if \( L \) intersects \( C \) at \( p_0 \) with multiplicity greater than one. We will write this in coordinates and derive an equation of the tangent line to \( C \) at \( p_0 \).

Let \( C \) be a projective curve in \( \mathbb{P}^2 \) given by a homogeneous polynomial \( F(x, y, z) \). Let \( p_0 \) have coordinates \((x_0 : y_0 : z_0)\). First, we will write the equation of a line \( L \) passing through \( p_0 \) and some other point \( p = (x : y : z) \) in parametric form:

\[
L = \{ p_0 + pt \mid t \in \mathbb{F} \} = \{(x_0 + xt, y_0 + yt, z_0 + zt) \mid t \in \mathbb{F} \}.
\]

Then the solutions to the equation \( F(p_0 + pt) = 0 \) produce the intersection points of \( C \) and \( L \).
Hence we find the tangent to \( F \) at \( p_0 \). This is, in fact, equivalent to the previous equation because of the Euler formula

\[
\nabla F(p_0) = \left( \frac{\partial F}{\partial x}(p_0), \frac{\partial F}{\partial y}(p_0), \frac{\partial F}{\partial z}(p_0) \right)
\]

is the gradient of \( F \) at \( p_0 \).

**Proof.** By linearity it is enough to prove the statement for any monomial \( F = x^iy^jz^k \). Write the expansions in \( t \)

\[
(x_0 + tx)^i = x_0^i + ix_0^{i-1}tx + \ldots, \quad (y_0 + ty)^j = y_0^j + jy_0^{j-1}ty + \ldots
\]

\[(z_0 + tz)^k = z_0^k + kz_0^{k-1}tz + \ldots
\]

Taking the product of the three expansions we obtain

\[
x_0^iy_0^jz_0^k + \left[ ix_0^{i-1}y_0^{j-1}z_0^k(x) + x_0^iy_0^{j-1}z_0^k(y) + x_0^iylkz_0^{k-1}(z) \right] t + \ldots
\]

It remains to notice that the first term is \( F(p_0) \) and the expression in the square brackets is the dot product of \( \nabla_{p_0} F \) and \( p = (x, y, z) \).

Now suppose \( p_0 \) lies on \( C \). Then \( F(p_0) = 0 \) and so \( t = 0 \) is a solution to \( F(p_0 + pt) = 0 \), i.e. \( L \) intersects \( C \) at \( p_0 \). Furthermore, from Lemma 2.47 we see that if \( \nabla_{p_0} F \cdot p = 0 \) then \( t = 0 \) is a multiple solution of \( F(p_0 + pt) = 0 \), i.e. \( L \) is tangent to \( C \) at \( p_0 \). We conclude

**Proposition 2.48.** Suppose a line \( L = \{p_0 + pt \mid t \in \mathbb{F} \} \) is tangent to \( C \) at \( p_0 \). Then \( F(p_0) = 0 \) and \( \nabla_{p_0} F \cdot p = 0 \). Consequently, \( \nabla_{p_0} F \cdot p = 0 \) is an equation of the tangent line to \( C \) at \( p_0 \).

Let us write the equation of the tangent line in coordinates:

\[
x \frac{\partial F}{\partial x}(p_0) + y \frac{\partial F}{\partial y}(p_0) + z \frac{\partial F}{\partial z}(p_0) = 0.
\]

**Remark 2.49.** You may have seen the following equation of the tangent line to \( F = 0 \) at \( p_0 \):

\[
\frac{\partial F}{\partial x}(p_0)(x - x_0) + \frac{\partial F}{\partial y}(p_0)(y - y_0) + \frac{\partial F}{\partial z}(p_0)(z - z_0) = 0.
\]

This is, in fact, equivalent to the previous equation because of the Euler formula (see Exercise 2.8): For any homogeneous polynomial \( F \) of degree \( n \)

\[
x \frac{\partial F}{\partial x} + y \frac{\partial F}{\partial y} + z \frac{\partial F}{\partial z} = nF.
\]

In our situation

\[
x_0 \frac{\partial F}{\partial x}(p_0) + y_0 \frac{\partial F}{\partial y}(p_0) + z_0 \frac{\partial F}{\partial z}(p_0) = nF(p_0) = 0.
\]

**Example 2.50.** Let \( C \) be a conic with equation \( F(x, y, z) = yz - x^2 = 0 \). Let us find the tangent to \( p_0 = (1 : 1 : 1) \). We have \( \nabla_{p_0} F = (-2y, z, y) \mid_{p_0} = (-2, 1, 1) \).

Hence \(-2x + y + z = 0 \) is the tangent line to \( C \) at \( (1 : 1 : 1) \).
2.3.2. Singular Points. Let $C$ be a projective curve, $C \subset \mathbb{P}^2$.

**Definition 2.51.** A point $p_0 \in C$ is called singular if $\nabla_{p_0}F = 0$. A point $p_0 \in C$ which is not singular is called smooth (or nonsingular, or regular).

Note that the tangent line is not defined at a singular point. In fact, any line passing through a singular point of $C$ satisfies the equation of the tangent line (see Proposition 2.48).

**Definition 2.52.** A curve $C$ without singular points is called smooth (or nonsingular).

**Example 2.53.** We find the singular points of a projective curve $C$ with equation $F(x, y, z) = 0$.

(a) Let $F(x, y, z) = y^2z - x^3 - x^2z$. We have $\nabla F = (-3x^2 - 2xz, 2yz, y^2 - x^2)$. Now if $\nabla F = 0$ then from the second component of $\nabla F$ we get $2yz = 0$, so either $y = 0$ or $z = 0$. In the first case $x^2 = 0$ from the third component of $\nabla F$, and $z$ is any non-zero, i.e. we obtain a singular point $(0 : 0 : 1)$. In the second case $x^2 = 0$ from the first component of $\nabla F$ and hence $y^2 = 0$ from the third component. This produces $(0, 0, 0)$ which does not correspond to a point in $\mathbb{P}^2$. Therefore, the only singular point of $C$ is $(0 : 0 : 1)$. The affine part of $C$ is the affine curve $y^2 - x^3 = 0$ which is called the nodal cubic. It is depicted in Figure 2.6.

![Figure 2.6. The nodal cubic $y^2 = x^3 + x^2$.](image-url)
(b) Let \( F(x, y, z) = y^2z - x^3 \). We have \( \nabla F = (-3x^2, 2yz, y^2) \). Again it is easy to see that \( \nabla F = 0 \) is satisfied only for \( x = 0, y = 0 \) and any \( z \). This defines the only singular point \( (0 : 0 : 1) \). The affine curve \( C_\Lambda \) is given by \( y^2 - x^3 = 0 \) and is called the cuspidal cubic, see Figure 2.7.

![Figure 2.7. The cuspidal cubic \( y^2 = x^3 \).](image)

Let \( C_\Lambda \) be the affine part of \( C \) defined by \( f(x, y) = F(x, y, 1) \). Similarly to the projective case we can show that the tangent line to \( C_\Lambda \) at \( p_0 = (x_0, y_0) \) has equation

\[
\frac{\partial f}{\partial x}(p_0)(x - x_0) + \frac{\partial f}{\partial y}(p_0)(y - y_0) = 0.
\]

Furthermore, \( p_0 \) is a singular point of \( C_\Lambda \) if and only if \( f(p_0) = 0 \) and \( \nabla f(p_0) = 0 \).

2.3.3. **Local description of singular points.** Let \( p_0 = (x_0 : y_0 : z_0) \) be a singular point of a curve \( C \subset \mathbb{P}^2 \). By changing the coordinate system we can ensure \( p_0 = (0 : 0 : 1) \). We are interested in the geometry of \( C \) near \( p_0 \), so we are going to consider the affine part of \( C \).

Let \( f(x, y) \) be the polynomial defining the affine part \( C_\Lambda \). We can write

\[
f(x, y) = a_{00} + (a_{10}x + a_{01}y) + (a_{20}x^2 + a_{11}xy + a_{02}y^2) + \cdots = f_0 + f_1(x, y) + f_2(x, y) + \cdots
\]

We call \( f_i(x, y) \) the **homogeneous part of \( f \) of degree \( i \)**. Now since \( (0, 0) \) lies on \( C_\Lambda \) we have \( f_0 = a_{00} = 0 \). Furthermore, since \( (0, 0) \) is a singular point of \( C_\Lambda \) we have \( \frac{\partial f}{\partial x}(0, 0) = \frac{\partial f}{\partial y}(0, 0) = 0 \), i.e. \( a_{10} = a_{01} = 0 \). In other words, \( (0, 0) \) is a singular point of \( C_\Lambda \) if and only if the homogeneous parts of \( f \) of degree 0 and 1 are identically zero.
DEFINITION 2.54. We say that $p_0 = (0, 0)$ is a multiple point of $C_A$ of order $k$ if $k$ is the smallest positive integer for which $f$ has a non-zero homogeneous part of degree $k$. A point of order 2 is called a double point, and a point of order 3 is called a triple point.

We remark that if $(0, 0)$ is a point of order 1 then it is a smooth point of the curve. In this case $f_1(x, y) = 0$ is the equation of the tangent line to the curve at $(0, 0)$.

EXAMPLE 2.55.

(a) The nodal cubic has equation $f(x, y) = x^2 - y^2 + x^3 = 0$. We see that $f_0 = f_1 = 0$, $f_2 = x^2 - y^2$, and $f_3 = x^3$. Therefore $(0, 0)$ is a double point.

(b) The cuspidal cubic has equation $f(x, y) = y^2 - x^3 = 0$, hence $f_0 = f_1 = 0$, $f_2 = y^2$, and $f_3 = -x^3$. Again $(0, 0)$ is a double point.

(c) Consider the affine curve with equation $y^2 = x^3 - x$. This time $f(x, y) = x + y^2 - x^3$, hence, $f_0 = 0$ and $f_1 = x$. Therefore, $(0, 0)$ is a smooth point of the curve. Also $x = 0$ is the equation of the tangent line to the curve at $(0, 0)$ (see Figure 2.8).

\[ \begin{align*} f_k(x, y) & = a_{k0}x^k + a_{k-1,1}x^{k-1}y + \cdots + a_{0k}y^k = a_{k0}(x - \alpha_1 y) \cdots (x - \alpha_k y), \end{align*} \]

\[ \begin{figure}[h] \centering \includegraphics[width=0.5\textwidth]{smooth_cubic.png} \caption{The smooth cubic $y^2 = x^3 - x$.} \end{figure} \]
where $\alpha_1, \ldots, \alpha_k$ are the roots of $a_k t^k + a_{k-1} t^{k-1} + \cdots + a_0$. The lines $x = \alpha_i y$, for $1 \leq i \leq k$ are the tangents in question.

**Example 2.56.**

(a) The nodal cubic has two branches with tangents given by $f_2 = x^2 - y^2 = (x - y)(x + y) = 0$, i.e. $x = y$ and $x = -y$ as Figure 2.6 confirms.

(b) For the cuspidal cubic $f_2 = y^2$, hence it has two branches with the same tangent $y = 0$, see Figure 2.7.

A singular curve with a triple point at $(0, 0)$ is given in Exercise 2.12.

### 2.4. Bézout’s Theorem and Applications

We already mentioned Bézout’s theorem (see the discussion at the end of Section 2.2.1) which counts the number of intersections of two projective curves over an algebraically closed field. In this section we will prove Bézout’s theorem and look at some applications such as the number of singular points, the number of double tangents, and rational curves. We will also prove three classical theorems: Pappus (around 400 AD), Pascal (1639), and Chasles (1885).

#### 2.4.1. Bézout’s Theorem.

The proof of Bézout’s theorem employs an important algebraic object, called the resultant. We will define it a little later, first we will need a lemma about univariate polynomials. As before, $\mathbb{K}$ denotes an algebraically closed field.

**Lemma 2.57.** $f, g \in \mathbb{K}[x]$ have a common root in $\mathbb{K}$ if and only if there exist non-zero $f_1, g_1 \in \mathbb{K}[x]$ such that $f_1 g_1 = g_1 f_1$ and $\deg f_1 < \deg f$, $\deg g_1 < \deg g$.

**Proof.** $(\Rightarrow)$ Suppose $f, g$ have a common root $\alpha \in \mathbb{K}$. Then $f = (x - \alpha) f_1$ and $g = (x - \alpha) g_1$ for some $f_1, g_1 \in \mathbb{K}$ with $\deg f_1 < \deg f$, $\deg g_1 < \deg g$. Also $f_1 g_1 = g_1 f_1$ from above.

$(\Leftarrow)$ Suppose $h = f_1 g_1 = g_1 f_1$ with $\deg f_1 < \deg f$, $\deg g_1 < \deg g$. Then the roots of $f$ and the roots of $g$ are also the roots of $h$. But $\deg h = \deg f + \deg g_1 < \deg f + \deg g$, hence the sets of the roots of $f$ and $g$ cannot be disjoint. $\square$

**Definition 2.58.** Let $f = a_0 + a_1 x + \cdots + a_n x^n$ and $g = b_0 + b_1 x + \cdots + b_m x^m$ be two polynomials in $\mathbb{K}[x, y]$. Then their resultant $R(f, g)$ the determinant of the following $(m + n) \times (m + n)$ matrix (assume the empty spaces are filled with 0’s)

\[
\begin{pmatrix}
  a_0 & a_1 & \ldots & a_n \\
  a_0 & a_1 & \ldots & a_n \\
  \vdots & \vdots & \ddots & \vdots \\
  b_0 & \ldots & b_m \\
  b_0 & \ldots & b_m \\
  \vdots & \vdots & \ddots & \vdots \\
  b_0 & \ldots & b_m \\
  b_0 & \ldots & b_m
\end{pmatrix}
\]
2. ALGEBRAIC CURVES

Example 2.59. Let \( f = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \) and \( g = b_0 + b_1 x + b_2 x^2 \).

\[
R(f, g) = \begin{vmatrix}
a_0 & a_1 & a_2 & a_3 & 0 \\
0 & a_0 & a_1 & a_2 & a_3 \\
b_0 & b_1 & b_2 & 0 & 0 \\
0 & b_0 & b_1 & b_2 & 0 \\
0 & 0 & b_0 & b_1 & b_2 \\
\end{vmatrix}
\]

\[
= a_0^2 b_2^3 + 3a_0 a_3 b_0 b_2 b_1 - 2a_0 a_2 b_0 b_2^2 - 2a_2 a_3 b_1 b_0^2 + a_0 a_1 b_2 b_1 + a_1 a_3 b_0 b_1^2 - a_2 a_3 b_0^2 b_2 - a_1 a_2 b_0 b_1 b_2 + a_0 a_2 b_1^2 b_2 - a_0 a_3 b_1^3 + a_1^2 b_0 b_2^2 - 2a_1 a_3 b_0 b_2^2 + a_3^2 b_0.
\]

Lemma 2.60. \( f, g \in \mathbb{K}[x] \) have a common root in \( \mathbb{K} \) if and only if \( R(f, g) = 0 \).

Proof. By Lemma 2.57, \( f, g \in \mathbb{K}[x] \) have a common root if and only if one can find non-zero \( f_1 \) and \( g_1 \) of degrees smaller than \( \deg f \) and \( \deg g \), respectively such that \( fg_1 = gf_1 \). Expanding both sides and comparing the coefficients we obtain a homogeneous linear system with the coefficients of \( g_1 \) and \( f_1 \) being the unknowns and the matrix being equivalent to the resultant matrix. The linear system has a non-trivial solution if and only if the determinant of the matrix, which is \( \pm R(f, g) \), is non-zero.

\[\square\]

Example 2.61. Let \( f = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \) and \( g = b_0 + b_1 x + b_2 x^2 \), as in Example 2.59. Put \( g_1 = c_0 + c_1 x \) and \( f_1 = d_0 + d_1 x + d_2 x^2 \). We want

\[
(a_0 + a_1 x + a_2 x^2 + a_3 x^3)(c_0 + c_1 x) = (b_0 + b_1 x + b_2 x^2)(d_0 + d_1 x + d_2 x^2).
\]

Comparing the coefficients we get a system

\[
\begin{align*}
1: & \quad a_0 c_0 = b_0 d_0 \\
x: & \quad a_1 c_0 + a_0 c_1 = b_1 d_0 + b_0 d_1 \\
x^2: & \quad a_2 c_0 + a_1 c_1 = b_2 d_0 + b_1 d_1 + b_0 d_2 \\
x^3: & \quad a_3 c_0 + a_2 c_1 = b_3 d_0 + b_2 d_1 + b_1 d_2 \\
x^4: & \quad a_3 c_1 = b_2 d_2
\end{align*}
\]

which in the matrix form looks like

\[
\begin{pmatrix}
a_0 & 0 & -b_0 & 0 & 0 \\
a_1 & a_0 & -b_1 & -b_0 & 0 \\
a_2 & a_1 & -b_2 & -b_1 & -b_0 \\
a_3 & a_2 & 0 & -b_2 & -b_1 \\
0 & a_3 & 0 & 0 & -b_2
\end{pmatrix}
\begin{pmatrix}
c_0 \\
c_1 \\
d_0 \\
d_1 \\
d_2
\end{pmatrix}
= \begin{pmatrix}
0 \\
0 \\
0 \\
0 \\
0
\end{pmatrix}.
\]

The determinant of the matrix equals \(-R(f, g)\). It is zero if and only if non-trivial such \( f_1 \) and \( g_1 \) exist.

Next, assume \( F(x, y, z) \) and \( G(x, y, z) \), treated as polynomials in \( x \), have degrees \( n \) and \( m \). (We can always make a change of coordinates so that \( F \) has a term \( a_n x^n \) and \( G \) has a term \( b_m x^m \), see Exercise 2.14.) In this case we can write

\[
F(x, y, z) = a_n x^n + a_{n-1}(y, z)x^{n-1} + \cdots + a_0(y, z),
\]

\[
G(x, y, z) = b_m x^m + b_{m-1}(y, z)x^{m-1} + \cdots + b_0(y, z),
\]

where \( a_i(y, z) \) and \( b_i(y, z) \) are homogeneous polynomials of degree \( n - i \) and \( m - i \), respectively. We have the following statement.

Lemma 2.62. The resultant \( R(F(x, y, z), G(x, y, z)) = R(y, z) \) is a homogeneous polynomial of degree \( mn \).
2.4. Bézout’s Theorem and Applications

Proof. By Exercise 2.7, it is enough to show that \( R(\lambda y, \lambda z) = \lambda^{nm} R(y, z) \).

First, let us look at the case \( n = 3, m = 2 \) as in Example 2.59. As we notice above \( a_i(\lambda y, \lambda z) = \lambda^{n-i}a_i(y, z) \) and \( b_i(\lambda y, \lambda z) = \lambda^{m-i}b_i(y, z) \), hence we have

\[
R(\lambda y, \lambda z) =
\begin{bmatrix}
\lambda^3 a_0 & \lambda^2 a_1 & \lambda a_2 & a_3 & 0 \\
0 & \lambda^3 a_0 & \lambda^2 a_1 & \lambda a_2 & a_3 \\
\lambda^2 b_0 & \lambda b_1 & b_2 & 0 & 0 \\
0 & \lambda^2 b_0 & \lambda b_1 & b_2 & 0 \\
0 & 0 & \lambda^2 b_0 & \lambda b_1 & b_2
\end{bmatrix}
\]

Now we are going to multiply the first two rows by \( \lambda \) and 1, and the last three rows by \( \lambda^2, \lambda, \) and 1, so the columns become rescaled by consecutive powers of \( \lambda \). We obtain

\[
\lambda^{1+0} \lambda^{2+1+0} R(\lambda y, \lambda z) =
\begin{bmatrix}
\lambda^4 a_0 & \lambda^3 a_1 & \lambda^2 a_2 & \lambda a_3 & 0 \\
0 & \lambda^3 a_0 & \lambda^2 a_1 & \lambda a_2 & a_3 \\
\lambda^2 b_0 & \lambda^2 b_1 & \lambda b_2 & 0 & 0 \\
0 & \lambda^2 b_0 & \lambda b_1 & b_2 & 0 \\
0 & 0 & \lambda^2 b_0 & \lambda b_1 & b_2
\end{bmatrix}
= \lambda^{4+3+2+1+0} R(y, z).
\]

Comparing the powers of \( \lambda \) on both sides we see that \( R(\lambda y, \lambda z) = \lambda^{6} R(y, z) \).

The general case goes along the same lines. The only thing we will check is that the powers of \( \lambda \) on both sides will give as the desired answer \( \lambda^{nm} \):

\[
\lambda^{(m-1)+\cdots+1+0} \lambda^{(n-1)+\cdots+1+0} R(\lambda y, \lambda z) = \lambda^{(n+m-1)+\cdots+1+0} R(y, z),
\]

so the power of \( \lambda \) on the right hand side is \( \frac{(n+m)(n+m-1)}{2} \) and the one on the left hand side is \( \frac{n(m-1)}{2} + \frac{n(n-1)}{2} \). You can check that their difference is \( nm \). \( \square \)

We want to define the intersection number \( (C \cdot E)_{p_0} \) of two curves \( C \) and \( E \) at a point \( p_0 \in \mathbb{P}^2 \). We assume that \( C \) and \( E \) have no common components and so the number of their intersection points is finite. Let \( F, G \in \mathbb{K}[x, y, z] \) be homogeneous polynomials which define \( C \) and \( E \) of degree \( n \) and \( m \), respectively. As before we assume that the coordinates are chosen such that the degree of \( F \) and \( G \) in \( x \) are also \( n \) and \( m \). Also we require that no two intersection points of \( C \) and \( E \) have the same \( x \)-coordinate, that is if \( (x : y : 1) \) and \( (x : y' : 1) \) lie in \( C \cap E \) then \( y = y' \); and similarly, if \( (x : y : 0) \) and \( (x : y' : 0) \) lie in \( C \cap E \) then \( y = y' \).

Definition 2.63. Let \( C, E \) be projective curves with equations \( F = 0 \) and \( G = 0 \) as above. Define the (local) intersection number \( (C \cdot E)_{p_0} \) at \( p_0 = (x_0 : y_0 : z_0) \) to be the multiplicity of \( (y_0 : z_0) \) as a root of the resultant \( R(F, G) \in \mathbb{K}[y, z] \). In other words, \( (C \cdot E)_{p_0} \) is the largest integer \( k \) such that \( (z_0y - y_0z)^k \) divides \( R(F, G) \).

Note that if \( p_0 \) is not a common point of \( C \) and \( E \) then \( (C \cdot E)_{p_0} = 0 \). If \( p_0 \) lies on a common component of \( C \) and \( E \) then \( (C \cdot E)_{p_0} \) is undefined as then the resultant \( R(F, G) \) is identically zero.

Remark 2.64. In the case of \( E \) being a straight line, we can give an alternative definition based on our discussion of tangent lines in Section 2.3.1. Recall that if \( E \) is given parametrically \( E = \{ (p_0 + pt) \mid t \in \mathbb{K} \} = \{ (x_0 + xt, y_0 + yt, z_0 + zt) \mid t \in \mathbb{K} \} \) and \( C \) has equation \( F(x, y, z) = 0 \) then \( E \) is tangent to \( C \) at \( p_0 \) if \( t = 0 \) is a multiple solution to \( F(p_0 + pt) = 0 \). Define \( (C \cdot E)_{p_0} \) to be the multiplicity of \( t = 0 \) as a root of \( F(p_0 + pt) \). In Exercise 2.13 you will check the equivalence of these definitions.
2.65. (Bézout’s theorem) Let $C$ and $E$ be two plane projective curves. Then either they have a common irreducible component or they intersect in exactly $\deg C \cdot \deg E$ number of points, counting with multiplicities. In other words,

$$\sum_{p \in \mathbb{P}^2} (C \cdot E)_p = \deg C \cdot \deg E.$$  

Proof. Let $F$ and $G$ be the homogeneous polynomials of degree $n = \deg C$ and $m = \deg E$ and the coordinates are chosen as above. Then a point $p_0 = (x_0 : y_0 : z_0)$ lies in $C \cap E$ if and only if $p_0$ is a common zero of $F(x, y, z)$ and $G(x, y, z)$ which, in turn, happens if and only if $(y_0 : z_0)$ is a root of the resultant $R(F, G)$. Also, by definition, the intersection multiplicity $(C \cdot E)_{p_0}$ equals the multiplicity of this root. By Lemma 2.62, and since $K$ is algebraically closed, the homogeneous polynomial $R(F, G) \in K[y, z]$ has exactly $nm$ roots, counting with multiplicities. If $C$ and $E$ have a common irreducible component then $R(F, G)$ is identically zero. □

2.4.2. The Veronese map. Let us start with a standard fact in linear algebra: For any two points in $\mathbb{R}^2$ there is a line containing them. More generally, for any $d$ points in $\mathbb{R}^d$ there is a hyperplane containing them. The same is true about points in the projective space. For any $d$ points in $\mathbb{P}^d$ there is a hyperplane $H = \{a_0x_0 + a_1x_1 + \cdots + a_dx_d = 0\}$ containing them. Here $(x_0 : \cdots : x_d)$ are the homogeneous coordinates in $\mathbb{P}^d$. Indeed, if $\{p_1, \ldots, p_d\}$ is a collection of $d$ points, $p_i = (p_{i0} : \cdots : p_{id}) \in \mathbb{P}^d$ then such $H$ containing them exists if and only if the system

$$\begin{align*}
a_0p_{10} + a_1p_{11} + \cdots + a_dp_{1d} &= 0 \\
a_0p_{20} + a_1p_{21} + \cdots + a_dp_{2d} &= 0 \\
& \vdots \\
a_0p_{d0} + a_1p_{d1} + \cdots + a_dp_{dd} &= 0
\end{align*}$$

has a non-trivial solution $(a_0, \ldots, a_d)$, which is always true since the number of variables is greater than the number of equations. Note that if the rank of the above matrix equals $d$ then there is a unique hyperplane $H$ containing the points $\{p_1, \ldots, p_d\}$. Similarly, if $d-1$ points $\{p_1, \ldots, p_{d-1}\}$ in $\mathbb{P}^d$, considered as vectors in $K^{d+1}$, are linearly independent then there is a 1-parameter family of hyperplanes in $\mathbb{P}^d$ containing these points.

Now the question we are interested in is: How many points in $\mathbb{P}^2$ determine a degree $n$ plane projective curve that contains them?

Example 2.66. (Conics). Let $C$ be a conic given by a degree 2 homogeneous polynomial $F(x, y, z) = a_{00}z^2 + a_{10}xz + a_{01}yz + a_{20}x^2 + a_{11}xy + a_{02}y^2$.

Now let $\{p_1, \ldots, p_5\}$ be five points in $\mathbb{P}^2$, then $C$ contains them if and only if $F(p_i) = 0$ for $1 \leq i \leq 5$. This is equivalent to a linear homogeneous system of 5 equations and 6 unknowns $(a_{00}, a_{10}, \ldots, a_{02})$, which always has a non-trivial solution. Therefore, any 5 points in $\mathbb{P}^2$ lie on a conic.

We can look at this example from a little bit more general point of view. Define the following map

$$\nu_2 : \mathbb{P}^2 \to \mathbb{P}^5, (x : y : z) \mapsto (z^2 : xz : yz : x^2 : xy : y^2).$$
Then five points \( \{ p_1, \ldots, p_5 \} \) in \( \mathbb{P}^2 \) will be mapped to five points in \( \mathbb{P}^5 \), namely \( \{ \nu_2(p_1), \ldots, \nu_2(p_5) \} \). Furthermore, \( \{ p_1, \ldots, p_5 \} \) lie on a conic with equation

\[
F(x, y, z) = a_{00}z^2 + a_{10}xz + a_{01}yz + a_{20}x^2 + a_{11}xy + a_{02}y^2 = 0
\]

if and only if the points \( \{ \nu_2(p_1), \ldots, \nu_2(p_5) \} \) lie on a hyperplane \( H \) defined by

\[
a_{00}x_0 + a_{10}x_1 + a_{01}x_2 + a_{20}x_3 + a_{11}x_4 + a_{02}x_5 = 0,
\]

where \( (x_0 : \cdots : x_5) \) are the homogeneous coordinates in \( \mathbb{P}^5 \).

**Definition 2.67.** Let \( d = \binom{n+2}{2} - 1 \). The map \( \nu_n : \mathbb{P}^2 \to \mathbb{P}^d \) defined by

\[
\nu_n : (x : y : z) \mapsto (z^n : xz^{n-1} : \cdots : xy^{n-1} : y^n)
\]

is called the \( n \)-th Veronese map. Here on the right hand side are all possible homogeneous monomials of degree \( n \) in \( x, y, z \).

We have the following property of the Veronese map.

**Proposition 2.68.** The map \( \nu_n : \mathbb{P}^2 \to \mathbb{P}^d \) is a well-defined one-to-one map.

**Proof.** To see that \( \nu_n \) is well-defined we need to check that if not all \( x, y, z \) are zero then not all \( (z^n : xz^{n-1} : \cdots : xy^{n-1} : y^n) \) are zero (should be clear), and that \( (x : y : z) \) and \( (\lambda x, \lambda y, \lambda z) \) are mapped to the same point in \( \mathbb{P}^d \). The latter is also clear since the components of the map all have the same homogeneous degree \( n \) in \( x, y, z \).

Now to show that \( \nu_n \) is one-to-one assume that \( \nu_n(x, y, z) = \nu_n(u, v, w) \). Then there exists \( \mu \in \mathbb{K}^* \) such that \( z^n = \mu w^n \), \( xz^{n-1} = \mu vw^{n-1} \), and so on up to \( y^n = \mu v^n \). In particular, \( x^n = \mu u^n \), \( y^n = \mu v^n \), and \( z^n = \mu w^n \) imply that \( x = \lambda_1 u \), \( y = \lambda_2 v \), and \( z = \lambda_3 w \) where \( \lambda_i \) are some \( n \)-th roots of \( \mu \). From here it is not hard to see that the other equations ensure that \( \lambda_1 = \lambda_2 = \lambda_3 \). \( \square \)

In the following theorem we generalize the result of Example 2.66.

**Theorem 2.69.** Let \( d = \binom{n+2}{2} - 1 \).

1. For any \( d \) points in \( \mathbb{P}^2 \) there is a degree \( n \) curve containing them.
2. For any \( d - 1 \) points in \( \mathbb{P}^2 \) there is a \( 1 \)-parameter family of degree \( n \) curves containing them.
3. There exist \( d + 1 \) points in \( \mathbb{P}^2 \) which do not lie on any curve of degree \( n \).

**Proof.** (1) Let \( \{ p_1, \ldots, p_d \} \) be \( d \) points in \( \mathbb{P}^2 \) and let \( \{ \nu_n(p_1), \ldots, \nu_n(p_d) \} \) be the corresponding images under the Veronese map. Then \( \{ p_1, \ldots, p_d \} \) lie on a degree \( n \) curve \( C \) if and only if \( \{ \nu_n(p_1), \ldots, \nu_n(p_d) \} \) lie on a hyperplane in \( \mathbb{P}^d \). Now the statement follows from the linear algebra discussion at the beginning of the subsection.

(2) This is similar to (1). All we need is to notice that there is a \( 1 \)-parameter family of hyperplanes containing \( d - 1 \) points \( \{ \nu_n(p_1), \ldots, \nu_n(p_{d-1}) \} \) in \( \mathbb{P}^d \).

(3) Let \( X = \nu_n(\mathbb{P}^2) \) be the image of the projective plane under the Veronese map. Clearly \( X \) does not lie in a hyperplane (otherwise there would be a degree \( n \) curve in \( \mathbb{P}^2 \) which contains all points of \( \mathbb{P}^2 \)). Therefore, one can choose \( d + 1 \) points \( q_1, \ldots, q_{d+1} \) in \( X \), which do not lie on a hyperplane. But then their preimages \( p_1, \ldots, p_{d+1} \), where \( p_i = \nu_n^{-1}(q_i) \), do not lie on any curve of degree \( n \) in \( \mathbb{P}^2 \). \( \square \)
Remark 2.70. We remark without proof that for almost every collection of \(d\) points in \(\mathbb{P}^2\) there is a unique curve of degree \(n\) that contains them. What we mean here is that if we were to pick a collection of \(d\) points at random then with probability one they lie on a unique curve of degree \(n\). A more precise definition requires the notion of the dimension of an algebraic set, which we don’t include here. This is similar to the fact that almost any collection of \(d\) vectors in \(\mathbb{R}^d\) are linearly independent.

Similarly, for almost every collection of \(d - 1\) points in \(\mathbb{P}^2\) there is a unique 1-parameter family of degree \(n\) curves that contain them.

2.4.3. Number of double points. As the first application of Bezout’s theorem we will see how many singular double points a plane projective curve of degree \(n\) may have.

We begin with a question from elementary plane geometry: Given \(n\) lines in the plane, what is the largest number of intersection points they may have? It is easy to see that for \(n = 1, 2, 3\) and 4 the answer is 0, 1, 3 and 6, respectively. Since adding a line to a collection of \(n - 1\) intersecting lines produces at most \(n - 1\) more intersection points, the answer is the triangular number \(\binom{n}{2}\). Translating this to the language of curves, we can say that a reducible curve of degree \(n\) may have up to \(\binom{n}{2}\) singular double points. For irreducible curves the answer is smaller.

Theorem 2.71. Let \(C\) be an irreducible plane projective curve of degree \(n\). Then it may have at most \(\binom{n-1}{2}\) double points.

Proof. Let \(C\) be an irreducible cubic. We have seen examples of irreducible cubics with one double point (Example 2.53). Suppose \(C\) has two double points \(p_1, p_2\). Then the line \(E\) containing them intersects \(C\) at two points with multiplicity two each, i.e. \((C \cdot E)_{p_1} = (C \cdot E)_{p_2} = 2\). But, by Bezout’s theorem, this is impossible as \(\deg C \cdot \deg E = 3\) and \(E\) cannot be a component of \(C\) since we assumed \(C\) to be irreducible.

The same idea works in general. Let \(N = \binom{n-1}{2}\) and assume \(C\) has \(N+1\) double points which we denote by \(p_1, \ldots, p_N, p_{N+1}\). Choose \(n - 3\) more points on \(C\), call them \(q_1, \ldots, q_n\). Now we have \(N+1+n-3 = \binom{n}{2} - 1\), hence, by Theorem 2.69 there is a curve \(E\) of degree \(n-2\) containing them. Let us now count the intersection number of \(C\) and \(E\). Each \(p_i\) comes with intersection number 2 and each \(q_i\) comes with intersection number at least 1, which gives the total of at least \(2(N+1) + n - 3 = n(n-2) + 1\). This contradicts Bezout’s theorem as \(\deg C = n\) and \(\deg E = n-2\). Also \(E\) cannot be a component of \(C\) because \(C\) is irreducible and has larger degree than \(E\).

We mention without proof that there do exist irreducible curves with exactly \(\binom{n-1}{2}\) double points, so the bound proved in the above theorem is sharp.

2.4.4. Rational Curves. Let us come back to the nodal cubic \(C\) with affine equation \(y^2 = x^3 + x^3\) from Example 2.53. Can we parametrize the points of \(C\) just like we parametrize all the points of a line? In other words, can we write the coordinates \((x, y) \in C\) as rational functions of a parameter \(t\)? The answer turns out to be yes, and Bezout’s theorem will help us here.

The idea is to consider the intersection of \(C\) with the lines through the origin. Since the origin is a double point of \(C\), every line \(L\) must intersect \(C\) exactly once
more with multiplicity one, by Bezout’s theorem. We can express the coordinates of this intersection point as polynomial functions of \( t \), the slope of \( L \).

Indeed, if \( L \) is given by \( y = tx \) then the intersections of \( C \) and \( L \) correspond to the solutions of \( t^2x^2 = x^3 + x^2 \), i.e. of \( x^2(x - (t^2 - 1)) = 0 \). Notice that \( x = 0 \) is a double root, that’s the multiplicity two intersection at the origin \((0, 0)\). The simple root \( x = t^2 - 1 \) produces the point \((t^2 - 1, t^3 - t)\) on \( C \). We thus obtain a rational (in this case polynomial) parametrization

\[
C = \{(t^2 - 1, t^3 - t) \mid t \in \mathbb{K}\}.
\]

Note that the values \( t = 1, -1 \) give us the same point on \( C \), the origin. The corresponding lines \( y = x, y = -x \) are the tangents to the two branches at the origin (see Figure 2.6). The intersection number of each of these lines with \( C \) at the origin is three. Summarizing, every \( t \in \mathbb{K} \) defines a unique point \((t^2 - 1, t^3 - t)\) on \( C \). Conversely, every point \((x, y)\) on \( C \), except the origin, corresponds to a unique value of \( t \). The origin corresponds to \( t = 1 \) and \( t = -1 \).

**Example 2.72.** Here is an example of a rational parametrization of a smooth conic (a unit circle), which you may have seen before. Let the conic \( C \) be given by \((x - 1)^2 + y^2 - 1 = 0\). As before take all lines \( y = tx \) through the origin. They intersect \( C \) at one more point: \((x - 1)^2 + t^2x^2 - 1 = 0\), i.e. \( x((1 + t^2)x - 2) = 0\). This gives us a rational parametrization

\[
C = \left\{ \left( \frac{2}{1+t^2}, \frac{2t}{1+t^2} \right) \mid t \in \mathbb{K} \right\}.
\]

Note that each \( t \in \mathbb{K} \) defines a unique point on \( C \) and every point \((x, y)\) in \( C \), except the origin, corresponds to a unique value of \( t \).

**Definition 2.73.** A rational parametrization of an affine curve \( C \subset \mathbb{A}^2 \) is a pair of rational functions \((x(t), y(t))\) such that all but finitely many points \((x, y)\) in \( C \) have the form \((x(t), y(t))\) for some \( t \in \mathbb{K} \).

**Definition 2.74.** A affine curve is called rational if it admits a rational parametrization. A projective curve is called rational if its affine part admits a rational parametrization.

The next theorem says that irreducible curves with the largest possible number of double points are, in fact, rational.

**Theorem 2.75.** Let \( C \) be an irreducible curve with \( \binom{n-1}{2} \) double points. Then \( C \) is rational.

**Proof.** Let \( p_1, \ldots, p_N \) be the double points, \( N = \binom{n-1}{2} \). As in the proof of Theorem 2.71 we fix other \( q_1, \ldots, q_{n-3} \) on \( C \). Since \( N + n - 3 = \binom{n}{2} - 2 \), by Theorem 2.69 there is a 1-parameter family of curves \( E_t \) of degree \( n - 2 \) containing \( \{p_1, \ldots, p_N, q_1, \ldots, q_{n-3}\} \). In other words, the coefficients of the polynomial defining \( E_t \) are linear functions in \( t \). By Bezout’s theorem, for every value of \( t \) (except for finitely many exceptions) the curve \( E_t \) intersects \( C \) at exactly one more point \( q \). The \((y, z)\)-coordinates of \( q \) can then be found from solving \( R(F, G) = 0 \) where \( F \) and \( G \) are the polynomials defining \( C \) and \( E \), respectively. They are rational functions in the coefficients of \( G \), and hence, rational functions of the parameter \( t \). \( \square \)

In Exercise 2.2 you will find a rational parametrization for the curve given by \((x^2 + y^2)^2 = (x^2 - y^2)z^2\).
2.4.5. Pappus, Pascal, and Chasles. In this section we will prove three classical theorems due to Pappus of Alexandria (around 400 AD), Pascal (1640) and Chasles (1885). In fact, the former two follow from the latter, so first, we are going to state Chasles’ theorem and use it to prove the other two, and then prove Chasles’ theorem after a bit of preparation.

**Theorem 2.76.** (Chasles) Let $C_1, C_2$ be plane projective cubics which intersect in 9 points. Then any cubic that contains any 8 of these points must contain all of them.

**Theorem 2.77.** (Pappus) Let $\{p_1 \overset{\sim}{\rightarrow} p_2 \overset{\sim}{\rightarrow} p_3\}$ and $\{q_1 \overset{\sim}{\rightarrow} q_2 \overset{\sim}{\rightarrow} q_3\}$ be two ordered triples of collinear points. Define $a_{ij}$ to be the intersection of lines $p_iq_j$ and $p_jq_i$ for every $1 \leq i < j \leq 3$. Then the three points $r_{12}, r_{13},$ and $r_{23}$ are collinear.

**Proof.** Let $C_1$ be the union of three lines $p_1q_2, p_2q_3,$ and $p_3q_1$. Similarly, let $C_2$ be the union of $p_1q_3, p_3q_2,$ and $p_2q_1$. Then $C_1$ and $C_2$ are two (reduced) cubic curves intersecting in 9 points:

$$\{p_1, p_2, p_3, q_1, q_2, q_3, r_{12}, r_{13}, r_{23}\}.$$  

Now let $N$ be the line containing $r_{12}, r_{13}$ and let $C$ be the union of three lines $L,$ $M$ and $N$ (see Figure 2.9). Then $C$ contains the first 8 points in the above list, and hence, must contain all of them, by Chasles’ theorem. Therefore $r_{23}$ also lies on $N$, which proves that $r_{12}, r_{13},$ and $r_{23}$ are collinear.

**Theorem 2.78.** (Pascal) Consider a hexagon inscribed into an irreducible conic $E$ in $\mathbb{P}^2$. Let $a_i$, for $1 \leq i \leq 6$, be the lines containing the sides of the hexagon. Then the intersection points of the lines containing opposite sides $p_{14} = a_1 \cap a_4$, $p_{25} = a_2 \cap a_5$, and $p_{36} = a_3 \cap a_6$ are collinear.

**Proof.** This time we let $C_1 = a_1 \cup a_3 \cup a_5$ and $C_2 = a_2 \cup a_4 \cup a_6$. These cubics intersect at the six vertices of the hexagon and at $p_{14}, p_{25}, p_{36}$ (see Figure 2.10).
Let $C$ be the union of the conic $E$ and the line $N$ containing $p_{14}$ and $p_{25}$. Since $C$ contains 8 of the intersection points of $C_1$ and $C_2$, it must also contain $p_{36}$, by Chasles’ theorem. Therefore, $p_{14}, p_{25}, p_{36}$ all lie on $N$, i.e. are collinear.

Before we give a proof of Chasles theorem we need to look closer at the Veronese map discussed in Section 2.4.2. Recall its definition:

$$
\nu_n : \mathbb{P}^2 \rightarrow \mathbb{P}^d, \quad \nu_n : (x : y : z) \mapsto (z^n : xz^n - 1 : \ldots : xy^{n-1} : y^n),
$$

where $d = \left(\frac{n+2}{2}\right) - 1$.

**Definition 2.79.** A set of points $\{p_1, \ldots, p_k\} \subset \mathbb{P}^2$ is said to impose independent conditions on curves of degree $n$ if its image $\{\nu_n(p_1), \ldots, \nu(p_k)\}$ is linearly independent (as a set of vectors in $\mathbb{A}^{d+1}$). Otherwise we say it fails to impose independent conditions on degree $n$ curves.

In Section 2.4.2 we saw that if $S = \{p_1, \ldots, p_d\}$ imposes independent conditions on degree $n$ curves then there is a unique degree $n$ curve $C$ containing $S$.

Here is an equivalent definition.

**Proposition 2.80.** $S = \{p_1, \ldots, p_k\}$ imposes independent conditions on degree $n$ curves if and only if for every subset $S'$ of size $|S'| = k - 1$ there exists a degree $n$ curve $C'$ containing $S'$, but not $S$.

**Proof.** By definition $S$ imposes independent conditions on degree $n$ curves if and only if the matrix $M = [\nu_n(p_1), \ldots, \nu(p_k)]$ has rank $k$ if and only if its columns span $\mathbb{A}^k$ if and only if for every $1 \leq i \leq k$ the equation $M(a_{00}, \ldots, a_{0n})^t = e_i$ has a solution (here $e_i$ is the $i$-th standard basis vector for $\mathbb{A}^k$). This is equivalent the existence of a hyperplane in $\mathbb{P}^d$ containing all of $\{\nu_n(p_1), \ldots, \nu(p_k)\}$, but $\nu(p_i)$, i.e. the existence of a degree $n$ curve containing $\{p_1, \ldots, p_k\}$, but $p_i$.

\[\square\]
Example 2.81.

(a) Any two points in \( \mathbb{P}^2 \) impose independent conditions on lines (degree 1 curves), since there is a line containing one point, but not the other.

(b) Any three points in \( \mathbb{P}^2 \) impose independent conditions on lines, unless they are collinear, since then there is a line containing any two of the points, but not the third.

(c) Any six points impose independent conditions on conics, unless they lie on a conic (note that a conic may be reduced, i.e. the union of two lines). Indeed, suppose they do not all lie on a conic. Then, as we saw in Example 2.66, any five of them will lie on a conic, but not the sixth.

More generally, we can say when a set does or does not impose independent conditions in some special cases.

Proposition 2.82. Any \( k \leq n + 1 \) points in \( \mathbb{P}^2 \) impose independent conditions on degree \( n \) curves.

Proof. Let \( S = \{ p_1, \ldots, p_k \} \). For each \( 1 \leq i \leq k - 1 \) choose a line \( L_i \) containing \( p_i \), but not \( p_k \). Let \( E \) be any curve of degree \( n - k + 1 \) not containing \( p_k \). Then the union \( C = L_1 \cup \cdots \cup L_{k-1} \cup E \) is a curve of degree \( n \) containing all of \( S \), but \( p_k \). By Proposition 2.80, \( S \) imposes independent conditions on degree \( n \) curves.

Proposition 2.83. Any \( n + 2 \) collinear points in \( \mathbb{P}^2 \) fail to impose independent conditions on degree \( n \) curves.

Proof. Let \( S \) be a set of \( n + 2 \) points contained in a line \( L \) and choose \( S' \subset S \) of size \( n + 1 \). By Bezout’s theorem any curve \( C \) of degree \( n \) containing \( S' \) must contain \( L \) as an irreducible component. Therefore, \( C \) must contain \( S \) as well. By Proposition 2.80, \( S \) fails to impose independent conditions on degree \( n \) curves.

Proposition 2.84. Any \( 2n + 2 \) points in \( \mathbb{P}^2 \) lying on a conic fail to impose independent conditions on degree \( n \) curves.

Proof. The proof is the same as for Proposition 2.83, except the line \( L \) must be replaced by a conic \( E \).

Now we can strengthen the statement of Proposition 2.82.

Theorem 2.85. Any \( k \leq 2n + 2 \) points in \( \mathbb{P}^2 \) impose independent conditions on degree \( n \) curves unless there are \( n + 2 \) that are collinear or \( k = 2n + 2 \) and they all lie on a conic.

Proof. Notice that the cases \( k \leq n + 1 \) and \( n = 1, 2 \) have been already considered in Proposition 2.82 and Example 2.81. Thus we may assume that \( k > n + 1 \) and \( n \geq 3 \). Suppose \( S = \{ p_1, \ldots, p_k \} \) fails to impose independent conditions on degree \( n \) curves. Then we will show that \( S \) falls into one of the two situations described in the statement of the theorem. In fact, it is enough to consider minimal such \( S \), i.e. whose proper subsets do impose independent conditions on degree \( n \) curves. Equivalently, any degree \( n \) curve containing all but one point of \( S \) must contain all of \( S \). We use induction on \( n \) and then \( k \). We have three cases.

Case 1. Suppose \( S \) contains \( n + 1 \) collinear points, \( S' \subset L \). Let \( S'' = S \setminus S' \) and so \( |S''| \leq n + 1 \). First, \( S'' \) must fail to impose independent conditions on degree \( n - 1 \) curves. Indeed, assuming the contrary there exists \( E \) of degree \( n - 1 \)
containing \( S'' \setminus \{ p_j \} \), but not \( p_j \), and hence \( L \cup E \) is a degree \( n \) curve containing \( S \setminus \{ p_j \} \), but not \( p_j \). This is a contradiction.

Now by the inductive hypothesis \( S'' \) must contain \( n+1 \) collinear points, \( S'' \subset N \). This implies that \( |S| = 2n + 2 \) and \( S \) is contained in the union of two lines \( L \cup N \).

**Case 2.** Suppose \( S \) contains \( I \geq 3 \) collinear points \( S' \subset L \). Then, as before, \( S'' = S \setminus S' \) must fail to impose independent conditions on degree \( n-1 \) curves. By induction, since \( |S''| \leq 2(n-1) + 2 \), it must contain \( n+1 \) collinear points and we are back in Case 1.

**Case 3.** Suppose no three points of \( S \) are collinear. Let \( \{ p_1, p_2, p_3 \} \) be any three points of \( S \). We let \( S'' = S \setminus \{ p_1, p_2, p_3 \} \) and \( S_i = S'' \cup \{ p_i \} \). Once again, each set \( S_i \) must fail to impose independent conditions on degree \( n-1 \) curves, and hence, by the inductive hypothesis, \( |S_i| = 2n \) and \( S_i \subset C_i \) for some conic \( C_i \). Then \( |S''| = 2n - 1 \), and so \( S'' \) contains at least 5 points (remember that \( n \geq 3 \)). By our assumption these 5 points impose independent conditions on conics, hence, are contained in a unique conic \( C \). Therefore, all \( C_i \) must coincide with \( C \), which implies that \( S \subset C \).

We are now ready to prove Chasles’ theorem. Let \( C_1 = \{ F_1 = 0 \} \) and \( C_2 = \{ F_2 = 0 \} \) be plane cubics intersecting in 9 points. Let \( S \) be any subset of 8 of these points. By Theorem 2.85, \( S \) imposes independent conditions on cubics, hence, there is a unique 1-parameter family of cubics containing \( S \) (see the proof of Theorem 2.69). Since \( C_1 \) and \( C_2 \) are two distinct cubics from this 1-parameter family, they span the entire family, i.e. any cubic \( C \) containing \( S \) has equation \( \lambda_1 F_1 + \lambda_2 F_2 = 0 \) for some \( \lambda_1, \lambda_2 \in \mathbb{K} \). Clearly, the ninth intersection point of \( C_1 \) and \( C_2 \) also satisfies this equation, i.e. all the nine points are contained in \( C \).

### 2.5. The Genus of a Curve

Recall from complex analysis how we compactify the complex plane \( \mathbb{C} \) by introducing one point at infinity \( z = \infty \). Topologically, \( \mathbb{C} \cup \{ \infty \} \) is a compact space homeomorphic to a sphere, and is called the Riemann sphere \( S^2 = \mathbb{C} \cup \{ \infty \} \). This is, in fact, the same as the complex projective line \( \mathbb{P}^1_{\mathbb{C}} \) we defined in Section 2.2.2

\[
\mathbb{P}^1_{\mathbb{C}} = \{ (x : y) \mid x, y \in \mathbb{C} \} = \{ (x : 1) \mid x \in \mathbb{C} \} \cup \{ (1 : 0) \} = \mathbb{C} \cup \{ \infty \} = S^2.
\]

Thus, we see that a complex projective line is topologically a real 2-dimensional sphere.

It turns out that every smooth complex projective curve is topologically a real orientable surface. This will allow us to define the genus of a projective curve (at least for curves over \( \mathbb{C} \)) as the genus of a real orientable surface. Geometry teaches us that every real orientable 2-dimensional manifold is homeomorphic to a sphere with \( g \) handles. The number of handles \( g \) is called the genus of a real orientable surface. For example, a sphere with one handle is a torus (the surface of a doughnut) and a sphere with two handles is the surface of a pretzel (see Figure 2.11). An ordinary sphere has no handles, hence it has genus zero.

We will start with an example.

**Example 2.86.** Let \( \lambda \in \mathbb{C} \) be a constant and \( \lambda \neq 0, 1 \). Let \( C \) be a cubic with equation \( y^2z = x(x-z)(x-\lambda z) \) in \( \mathbb{P}^2_{\mathbb{C}} \). It has complex dimension 1, which means that a neighborhood of every point in \( C \) “looks like” a neighborhood of a point
in \( \mathbb{C} \). Thus, \( C \) has real dimension 2, i.e. is a real surface. In fact, it is a compact orientable surface embedded into \( \mathbb{P}^2_{\mathbb{C}} \). We consider the projection

\[
\pi : \mathbb{P}^2_{\mathbb{C}} \to \mathbb{P}^1_{\mathbb{C}}, \quad (x : y : z) \mapsto (x : z).
\]

Note that it is undefined at \((0 : 1 : 0)\), so we set \((0 : 1 : 0) \mapsto (1 : 0) \in \mathbb{P}^1_{\mathbb{C}}\). In Figure 2.12 we show the real part of \( C \) and how it is projected on the \( x \)-axis. Note that every point \( p \in \mathbb{P}^1_{\mathbb{C}} \) has exactly two preimages on \( C \) unless \( p \) lies in the set of four exceptional points \( \Sigma = \{(\lambda : 1), (0 : 1), (1 : 1), (1 : 0)\} \). (The only preimage of \((1 : 0)\) is the point \((0 : 1 : 0) \in C\).) We will denote the four points by \( \lambda, 0, 1, \) and \( \infty \). In other words, \( \pi \) is a double covering of \( \mathbb{P}^1_{\mathbb{C}} \setminus \Sigma \) by two parts of \( C \), each looks like \( \mathbb{P}^1_{\mathbb{C}} \) with four points removed. Analytically, they are the two sheets of the 2-valued function \( y = \sqrt{x(x - 1)(x - \lambda)} \). We sketch the process of gluing the two sheets in Figures 2.13–2.15.

\[\text{Figure 2.11. Two surfaces of genus 1 and 2.}\]

\[\text{Figure 2.12. The projection of a cubic onto } \mathbb{P}^1_{\mathbb{C}}.\]
Each sheet is a sphere without the four points in $\Sigma$. The spheres are glued along the four cuts to form a surface homeomorphic to a torus. Therefore the cubic $C$ has genus 1.
The following theorem describes the basic idea for constructing Riemann surfaces from a smooth complex projective curve.

**Theorem 2.87.** Let \( C \subset \mathbb{P}^2 \) be a smooth plane projective curve and \( \pi : \mathbb{P}^2 \to \mathbb{P}^1, \ (x:y:z) \mapsto (x:z) \) be the vertical projection. Then every point \( p \in \mathbb{P}^1 \) has \( n = \deg C \) preimages under \( \pi \), unless \( p \in \Sigma \), the projection of points in \( C \) with a vertical tangent. Moreover, \( \mathbb{P}^1 \setminus \Sigma \) is covered by \( n \) sheets homeomorphic to a sphere with a finite number of points excluded. These sheets are glued to form a compact real orientable surface, the Riemann surface of \( C \).

The points in \( \Sigma \) have a special name.

**Definition 2.88.** Points \( p \in \mathbb{P}^1 \) with \( |\pi^{-1}(p) \cap C| < \deg C \) are called the ramification points of the map \( \pi|_C \). The map \( \pi|_C : C \to \mathbb{P}^1 \) is called a ramified \( n \)-covering.

### 2.5.1. The Euler Characteristic

We have already mentioned that every real compact orientable surface is topologically a sphere with \( g \) handles, where \( g \) is called the genus of the surface. Now we will define a related notion of the Euler characteristic. It can be defined combinatorially via triangulations.

Let \( S \) be a real compact orientable surface. A **triangulation** of \( S \) is a subdivision of \( S \) into finitely many triangles \( S = T_1 \cup \cdots \cup T_k \) in such a way that every pair of triangles either is disjoint or intersects in a common side or a common vertex. The triangles are called the **faces** and the sides are called the **edges** of the triangulation.

**Definition 2.89.** Let \( V, E, \) and \( F \) denote the number of vertices, edges, and faces in a triangulation of \( S \). The **Euler characteristic** \( \chi(S) \) is the number \( V - E + F \).

It turns out that \( \chi(S) \) is independent of the choice of a triangulation and is a topological invariant of the surface.

**Example 2.90.**

(a) Let \( S^2 \) be a sphere. We can construct a triangulation of \( S^2 \) by starting with a triangular pyramid (a 3-simplex) and blowing it up like a balloon. We obtain a sphere with a triangulation corresponding to the boundary of the simplex. We have \( V = 4, E = 6, \) and \( F = 4 \), so \( \chi(S^2) = 4 - 6 + 4 = 2 \).

(b) Let \( D^2 \) be a disc. The simplest triangulation we can choose is the one consisting of one triangle and three edges on the boundary of \( D \). Its Euler characteristic is \( \chi(D) = 3 - 3 + 1 = 1 \). Also the boundary \( \partial D \) of \( D \) has the Euler characteristic \( \chi(\partial D) = 3 - 3 + 0 = 0 \).
You should check that the Euler characteristic has the following property.

**Proposition 2.91.** The Euler characteristic is additive, i.e.

\[ \chi(S_1 \cup S_2) = \chi(S_1) + \chi(S_2) - \chi(S_1 \cap S_2). \]

The following theorem establishes a relation between the genus and the Euler characteristic.

**Theorem 2.92.** Let \( S \) be a compact orientable surface of genus \( g \). Then

\[ \chi(S) = 2 - 2g. \]

**Proof.** The proof is by induction on \( g \). If \( g = 0 \) then \( S \cong \mathbb{S}^2 \) and \( \chi(\mathbb{S}^2) = 2 \) as above. Let \( S_{g-1} \) be a surface of genus \( g - 1 \), and so \( \chi(S_{g-1}) = 2 - 2(g - 1) = 4 - 2g \), by the inductive hypothesis. We can construct a surface of genus \( g \) by attaching a handle to \( S_{g-1} \). Namely, let \( U \) be \( S_{g-1} \) with two disjoint discs removed. Take a “handle” \( V \), i.e. a sphere with two disjoint discs removed, and glue the handle \( V \) to \( U \) along the boundaries of the two pairs of removed discs, as in Figure 2.17. The resulting surface \( S_g = U \cup V \) has genus \( g \). To compute the Euler characteristic

\[
\text{Figure 2.17. Attaching a handle to } S_{g-1}.
\]

we will use the additivity property. First, \( \chi(V) = 2 - 1 - 1 = 0 \), as the Euler characteristic of the sphere is 2 and the Euler characteristic of each disc is 1. Note that the discs intersect the sphere in circles which have zero Euler characteristic. Similarly,

\[ \chi(U) = \chi(S_{g-1}) - 1 - 1 = 4 - 2g - 1 - 1 = 2 - 2g. \]

Therefore

\[ \chi(S_g) = \chi(U) + \chi(V) - \chi(U \cap V) = 2 - 2g + 0 - 0 = 2 - 2g, \]

where again we used the fact that \( U \cap V \) is a disjoint union of two circles, hence has zero Euler characteristic. The formula is proved. \( \square \)

**2.5.2. The Riemann–Hurwitz Formula.** Our goal now is to see how the Euler characteristic behaves when one has a ramified covering \( \pi: S_2 \to S_1 \). First, we consider \( n \)-coverings.

**Definition 2.93.** Let \( S_1, S_2 \) be two surfaces. A map \( \pi: S_2 \to S_1 \) is called an \( n \)-covering if every point \( p \in S_1 \) has a neighborhood \( U \subset S_1 \) such that \( \pi^{-1}(U) \) is homeomorphic to a disjoint union of \( n \) copies of \( U \).
Theorem 2.94. Let \( \pi : S_2 \to S_1 \) be an \( n \)-covering. Then \( \chi(S_2) = n \chi(S_1) \).

Proof. Choose a fine enough triangulations of \( S_1 \) and \( S_2 \) such that for any triangle \( \Delta \) in the triangulation of \( S_1 \) the preimage \( \pi^{-1}(\Delta) \) is a union of disjoint triangles in the triangulation of \( S_2 \). Then if \( \chi(S_1) = V - E + F \) then \( \chi(S_2) = nV - nE + nF = \chi(S_1) \). \( \square \)

As an immediate consequence of Theorem 2.92 and Theorem 2.94 we obtain:

Corollary 2.95. Let \( S_1 \) and \( S_2 \) be compact orientable surfaces of genera \( g_1 \) and \( g_2 \). If \( \pi : S_2 \to S_1 \) is a covering then \( g_1 - 1 \) divides \( g_2 - 1 \).

The following theorem, which we refer to as the Riemann–Hurwitz formula relates the Euler characteristics for a ramified covering of two surfaces.

Theorem 2.96. (The Riemann–Hurwitz formula) Consider a ramified \( n \)-covering \( \pi : S_2 \to S_1 \) with ramification points \( \{p_1, \ldots, p_k\} \subset S_1 \). Let \( n_i \) be the number of preimages of \( p_i \) under \( \pi \). Then

\[
\chi(S_2) = n \left( \chi(S_1) - k \right) + \sum_{i=1}^{k} n_i = n \chi(S_1) - \sum_{i=1}^{k} (n - n_i).
\]

Proof. Choose a triangulation of \( S_1 \) which contains small enough triangles \( \Delta_1, \ldots, \Delta_k \) such that they are disjoint and \( p_i \in \Delta_i \) for \( 1 \leq i \leq k \). Put \( \Sigma_1 = \Delta_1 \cup \cdots \cup \Delta_k \) and \( S_1 = T_1 \cup \Sigma_1 \), where \( T_1 \) is the union of the remaining triangles in the triangulation of \( S_1 \). Then

\[
(2.5) \quad \chi(S_1) = \chi(T_1) + \chi(\Sigma_1) - \chi(T_1 \cap \Sigma_1) = \chi(T_1) + k,
\]

since \( \Sigma_1 \) consists of \( k \) disjoint discs, each of Euler characteristic one, and \( T_1 \cap \Sigma_1 \) is a union of disjoint circles, each of Euler characteristic zero.

Now consider the triangulation of \( S_2 \) obtained by taking the preimages: \( S_2 = T_2 \cup \Sigma_2 \), for \( T_2 = \pi^{-1}(T_1) \) and \( \Sigma_2 = \pi^{-1}(\Sigma_1) \). Then

\[
(2.6) \quad \chi(S_2) = \chi(T_2) + \chi(\Sigma_2) - \chi(T_2 \cap \Sigma_2) = \chi(T_2) + n_1 + \cdots + n_k,
\]

since there are precisely \( n_i \) triangles in the preimage of \( \Delta_i \) and \( T_2 \cap \Sigma_2 \) is still a union of disjoint circles. It remains to notice that \( \pi : T_2 \to T_1 \) is an \( n \)-covering, so by Theorem 2.94, \( \chi(T_2) = n \chi(T_1) \), and the statement follows from (2.5) and (2.6). \( \square \)

2.5.3. The Plücker Formula. As an application of the Riemann–Hurwitz formula we will prove the Plücker formula which computes the genus of a smooth plane projective curve. We will start with an example.

Example 2.97. Let \( C \) be defined by \( x^n + y^n + z^n = 0 \). You should check that this is a smooth curve. The projection \( \pi : C \to \mathbb{P}_C^1 \), \( (x : y : z) \mapsto (x : z) \) is well-defined as \((0 : 1 : 0)\) does not lie on \( C \). Now take any point \( p = (x_0 : z_0) \in \mathbb{P}_C^1 \). What is \( \pi^{-1}(p) \)? We have

\[
\pi^{-1}(p) = \{(x_0 : y : z_0) \mid x_0^n + y^n + z_0^n = 0 \}.
\]

In other words, \( \pi^{-1}(p) \) consists of points whose \( y \)-coordinate is a solution to \( y^n = -(x_0^n + z_0^n) \). This equation has \( n \) distinct solutions in \( \mathbb{C} \) unless \( x_0^n + z_0^n = 0 \) in which case there is only one solution. Thus the points \( p = (x_0 : z_0) \) with \( x_0^n + z_0^n = 0 \) are the ramification points with exactly 1 preimage. We can write them explicitly. Indeed, let \( \xi_1, \ldots, \xi_n \) be the \( n \)-th roots of \(-1 \). Then the ramification points are
2.5. THE GENUS OF A CURVE

\( (\xi_1 : 1), \ldots, (\xi_n : 1) \). (We should point out that \( x_0 = y_0 = 0 \), although satisfies the above equation, does not define a point in \( \mathbb{P}^1_C \).

Now we can apply the Riemann–Hurwitz formula with \( S_2 = C, S_1 = \mathbb{P}^1_C, k = n, \) and \( n_i = 1 \) for \( 1 \leq i \leq n \):

\[
\chi(C) = n \chi(\mathbb{P}^1_C) - \sum_{i=1}^{n} (n - 1) = 2n - n(n - 1) = 3n - n^2.
\]

On the other hand, \( \chi(C) = 2 - 2g \), where \( g \) is the genus of \( C \). Therefore,

\[
g = \frac{2 - \chi(C)}{2} = \frac{n^2 - 3n + 2}{2} = \frac{(n - 1)(n - 2)}{2}.
\]

It turns out that the same formula holds for any smooth plane projective curve.

**Theorem 2.98. (The Plücker Formula)** Let \( C \) be a smooth plane projective curve of degree \( n \). Then the genus of \( C \) equals \( \frac{(n-1)(n-2)}{2} \).

**Proof.** Let \( C \) be defined by \( F(x, y, z) = 0 \). We can choose coordinates such that \( F \) has degree \( n \) in \( y \) and such that \( (0 : 1 : 0) \) does not lie on \( C \) and so the projection \( \pi : C \to \mathbb{P}^1_C, (x : y : z) \mapsto (x : z) \) is well-defined. Now for every \( p = (x_0 : z_0) \in \mathbb{P}^1_C \) the preimage \( \pi^{-1}(p) \) consists of points \( (x_0 : y : z_0) \) on \( C \) whose \( y \)-coordinates are the roots of \( F(x_0, y, z_0) = 0 \). Since \( F \) has degree \( n \) in \( y \), the number of preimages \( |\pi^{-1}(p)| \) is \( n \) unless \( y \) is a multiple root of \( F(x_0, y, z_0) \). The latter means that the vertical line \( L_p \) through \( p \) is tangent to \( C \). Points on \( C \) with a vertical tangent satisfy \( \frac{\partial F}{\partial y} = 0 \), i.e. are the intersection points of \( C \) and \( E = \{ \frac{\partial F}{\partial y} = 0 \} \). By Bezout’s theorem their number is \( \deg C \cdot \deg E = n(n - 1) \). Again, by changing coordinates if needed, we may assume that the points with a vertical tangent have different \( y \)-coordinates, i.e. they project to exactly \( k = n(n - 1) \) points \( \{p_1, \ldots, p_k\} \) in \( \mathbb{P}^1_C \). These are precisely the ramification points of \( \pi \). Also \( n_i = |\pi^{-1}(p_i)| = n - 1 \). Therefore, by the Riemann–Hurwitz formula

\[
\chi(C) = 2n - \sum_{i=1}^{k} (n - n_i) = 2n - n(n - 1) = 3n - n^2,
\]

\begin{figure}[h]
\centering
\includegraphics[scale=0.5]{figure2.18.png}
\caption{\( p_1 \) is a ramification point of \( \pi : C \to \mathbb{P}^1_C \).}
\end{figure}
which implies that
\[ g = \frac{2 - \chi(C)}{2} = \frac{n^2 - 3n + 2}{2} = \frac{(n - 1)(n - 2)}{2}. \]

### 2.5.4. Hyperelliptic curves.

Now we will look at a special class of plane curves called hyperelliptic. They have a singular point, so we need first to define the genus of a singular curve.

**Definition 2.99.** Let \( \Sigma \) be the set of singular points of \( C \). Then \( C \setminus \Sigma \) is a real orientable surface homeomorphic to a sphere with \( g \) handles with a finite number of points removed. The number \( g \) is called the **genus** of \( C \).

Let \( f \in \mathbb{K}[x] \) be a polynomial of degree \( n \geq 3 \) with distinct roots. It defines a projective curve \( C \) of degree \( n \) whose affine part has equation \( y^2 = f(x) \). If \( n = 3 \) or \( n = 4 \) the curve \( C \) is called **elliptic**; for \( n > 4 \) it is called **hyperelliptic**.

First, let us look at the case of an elliptic curve with \( n = 3 \). The homogeneous equation of \( C \) is \( y^2z = x^3 + a_2x^2z + a_1xz^2 + a_0z^3 \). You should check that \( C \) is smooth as long as \( f(x) = x^3 + a_2x^2 + a_1x + a_0 \) has no multiple roots. By the Plücker formula the genus of \( C \) equals \( (3-1)(3-2)/2 = 1 \). A particular case of this (which is, in fact, equivalent to the general situation) was considered in Example 2.86.

Now let \( f(x) = x^n + \cdots + a_1x + a_0 \), \( n \geq 4 \), and let \( C \) be defined by the homogeneous polynomial
\[ F(x,y,z) = y^2z^{n-2} - (x^n + \cdots + a_1xz^{n-1} + a_0z^n). \]

The gradient of \( F \) equals
\[ (-nx^{n-1} + \cdots + a_1z^{n-1})2yz^{n-2}, (n-2)y^2z^{n-3} - (a_{n-1}x^{n-1} + \cdots + na_0z^{n-1}) \].

In the affine part \( z = 1 \) the singular points on \( C \) are defined by \( y = 0 \) and \( f'(x) = 0 \).

In other words, they are the common roots of \( f \) and \( f' \), which there are none, since we assumed that \( f \) has no multiple roots. However, when \( z = 0 \), the gradient becomes \( (-nx^{n-1}, 0, -a_{n-1}x^{n-1}) \) which equals zero at \( (0 : 1 : 0) \). Thus \( (0 : 1 : 0) \) is the only singularity on \( C \). Notice also that this is the only point of \( C \) on the infinite line \( z = 0 \).

To compute the genus of \( C \) we turn back to the Riemann–Hurwitz formula. Consider the vertical projection \( \pi : C \to \mathbb{P}_C^1 \) defined by \( (x : y : z) \mapsto (x : z) \) and \( (0 : 1 : 0) \mapsto (1 : 0) \). For every point \( p = (x_0 : 1) \in \mathbb{P}_C^1 \) the preimages in \( C \) are points \( (x_0 : y : 1) \) satisfying the quadratic equation \( y^2 = f(x_0) \). Hence \( |\pi^{-1}(p)| = 2 \), unless \( x_0 \) is a root of \( f \), in which case \( |\pi^{-1}(p)| = 1 \).

Therefore \( \pi \) is a double covering with \( n \) ramification points corresponding to the roots of \( f \), and possibly the point \( (1 : 0) \). So let us see how the two sheets are glued in a neighborhood of \( (0 : 1 : 0) \). If \( (1 : 0) \) is not a ramification point then when we go around a small circle centered at \( (1 : 0) \) the preimages stay on one sheet and go around a small circle centered at \( (0 : 1 : 0) \). If \( (1 : 0) \) is a ramification point then the preimages will start on one sheet and end up on the other. Let \( (1 : z) \) with \( |z| = \rho \) be the points on the circle of radius \( \rho \), centered at \( (1 : 0) \). The preimages of \( (1 : z) \) satisfy
\[ y^2z^{n-2} = 1 + \cdots + a_1z^{n-1} + a_0z^n, \]
or equivalently
\[ y^2 = z^2 - n + \cdots + a_1z + a_0z^2. \]
Since $|z|$ is small, the main term in the right hand side is $z^{2-n}$ and we can approximate $C$ near $(0 : 1 : 0)$ by the curve with equation
$$y^2 = z^{2-n}.$$ We can parametrize the circle as $z = \rho e^{i\theta}$ for $0 \leq \theta \leq 2\pi$. Then the above equation becomes
$$y^2 = \rho^{2-n} e^{i\theta(2-n)}.$$ If $n = 2m$ then we have two sheets
$$y = \pm \rho^{1-m} e^{i\theta(1-m)}$$ and the preimage of the circle $|z| = \rho$ stays on one sheet. If $n = 2m - 1$ then
$$y = \pm \rho^{3/2-m} e^{i\theta(1-m)} e^{i\theta/2}$$ and the preimage of the circle $|z| = \rho$ switches from one sheet to the other.

Finally we can apply the Riemann–Hurwitz formula. If $n = 2m$ then the projection $\pi : C \setminus \{(0 : 1 : 0)\} \to \mathbb{P}^1_C$ is a double covering with $k = n$ ramification points and each $n_i = 1$. Thus
$$2 - 2g = \chi(C) = 2\chi(\mathbb{P}^1_C) - n = 4 - 2m$$ and so $g = m - 1 = \frac{n-2}{2}$. If $n = 2m - 1$ then there are $k = n + 1$ ramification points with $n_i = 1$. Hence
$$2 - 2g = \chi(C) = 2\chi(\mathbb{P}^1_C) - (n + 1) = 4 - 2m$$ and so $g = m - 1 = \frac{n-1}{2}$. We have proved the following result.

**Proposition 2.100.** The genus of a hyperelliptic curve of degree $n$ equals $\left\lfloor \frac{n-1}{2} \right\rfloor$.

**Exercises**

**Exercise 2.1.** Construct a field of 8 elements by choosing an irreducible polynomial over $\mathbb{Z}_2$. Find its primitive elements.

**Exercise 2.2.** List all monic irreducible polynomials of degrees 1, 2, and 4 over $\mathbb{Z}_2$. Show that their product equals $x^{16} - x$.

**Exercise 2.3.** Show that if $F$ is algebraically closed then $F$ is infinite. (Hint: Assuming $F$ is finite construct a polynomial over $F$ which is not equal to zero at any of the elements of $F$.)

**Exercise 2.4.** Consider ideals $I = \langle x, y \rangle$ and $J = \langle ax + by, cx + dy \rangle$ in $\mathbb{Q}[x, y]$, where $a, b, c, d$ are some constants. Find a necessary condition on these constants that guarantees $I = J$. (Hint: First show that $I \subseteq J$ if the generators of $I$ lie in $J$.)

**Exercise 2.5.** Factor $x^3 - y^3$ into a product of irreducible polynomials in $\mathbb{Q}[x, y]$ and in $\mathbb{C}[x, y]$. Don’t forget to show the irreducibility of the factors.

**Exercise 2.6.** Prove that $x^{p^n} - x$ equals the product of all monic irreducible polynomials over $\mathbb{F}_p$ whose degree divides $n$. (Hint: ??????)
Exercise 2.7. Show that a polynomial \( F \in K[x, y, z] \) is homogeneous of degree \( n \) if and only if it satisfies \( F(\lambda x, \lambda y, \lambda z) = \lambda^n F(x, y, z) \) for any \( \lambda \in K \).

Exercise 2.8. Let \( F \in K[x, y, z] \) be homogeneous of degree \( n \). Prove
\[
x \frac{\partial F}{\partial x} + y \frac{\partial F}{\partial y} + z \frac{\partial F}{\partial z} = nF.
\]

Exercise 2.9. Let \( M \in \text{GL}(3, K) \) be an invertible \( 3 \times 3 \) matrix over a field \( K \). It defines an invertible linear transformation on the set of lines in \( \mathbb{A}^3 \) passing through the origin, i.e on \( \mathbb{P}^2 \). Show that two matrices \( M \leftrightarrow M' \) define the same transformation on \( \mathbb{P}^2 \) if and only if \( M' = (\lambda I)M \), where \( \lambda I \) is a scalar matrix. Deduce from this that the group of invertible linear transformations of \( \mathbb{P}^2 \), denoted by PGL(3, K), is isomorphic to \( \text{GL}(3, K)/Z \), where \( Z \) is the subgroup of scalar matrices (the center).

Exercise 2.10. Find the intersection of the projective curve given by \( y^2z = x^3 - xz^2 \) and the lines (a) \( x = y \); (b) \( x = z \); (c) \( y = z \) (you don’t have to write down the coordinates in part (c), but give a description in terms of the number of real/complex points in the affine part and at infinity). Are there multiple intersection points? Sketch the affine part of the curve and the lines in \( \mathbb{R}^2 \).

Exercise 2.11. Find the singular points of the projective curve \( y^2z^2 = x(x^2 - z^2)(x - 2z) \). Find all the (smooth) points of the curve where the tangent is vertical (note that the tangent is vertical if and only if \( F_y = 0 \)).

Exercise 2.12. Show that \((0, 0)\) is a triple point of the affine curve \((x^2+y^2)^2 + 3x^2y - y^3 = 0\). What are the three tangents to the curve at the origin?

Exercise 2.13. Let \( C \) be a curve with equation \( F(x, y, z) = 0 \) and consider a line \( E \) given parametrically by
\[
E = \{(p_0 + pt \mid t \in K) = \{(x_0 + xt, y_0 + yt, z_0 + zt) \mid t \in K \}.
\]
Define \((C \cdot E)_{p_0}\) to be the multiplicity of \( t = 0 \) as a root of \( F(p_0 + pt) \). Show that \((C \cdot E)_{p_0}\) coincides with the local intersection number as defined in Definition 2.63. (Hint: ? ? ? ? ? ? ? ?)

Exercise 2.14. Let \( F(x, y, z) \) be a homogeneous polynomial of degree \( n \). Show that one can make a linear change of variables \( y \mapsto x + \lambda_1 y, z \mapsto x + \lambda_2 x \), so the resulting polynomial has degree \( n \) in \( x \).

Exercise 2.15. Prove that for any four points in \( \mathbb{P}^2 \), no three of which are collinear, there exists a linear transformation of \( \mathbb{P}^2 \) which sends them to \((1:0:0),(0:1:0),(0:0:1), \) and \((1:1:1)\).

Exercise 2.16. Find a rational parametrization of the curve \( x^2 - y^2 = 1 \). (Hint: Consider lines through \((-1,0)\).)
EXERCISE 2.17. Find a rational parametrization of the curve \((x^2 - y^2)^2 = y^3\).
(Hint: It has a singular point which is a triple point.)

EXERCISE 2.18. Find all conics containing the five points \((0 : 0 : 1), (1 : 0 : 1), (0 : 1 : 1), (1 : 1 : 0),\) and \((-1 : 1 : 1)\). Do these points impose independent conditions on conics?

EXERCISE 2.19. Find all conics containing the four points \((0 : 0 : 1), (0 : 1 : 1),\) and \((\pm 1 : 0 : 1)\). Do these points impose independent conditions on conics?

EXERCISE 2.20. Find all conics containing the three points \((0 : 0 : 1), (\pm i : 1 : 0)\) and which are tangent to the line \(x + y = 0\) at \((0 : 0 : 1)\).

EXERCISE 2.21. Let \(S\) consist of the integer points of the square \([0, 2] \times [0, 2]\) in \(\mathbb{R}^2\). Does \(S\) impose independent conditions on the cubics? Explain.

EXERCISE 2.22. Find a rational parametrization of the degree 4 curve \(C\) given by equation \((x^2 + y^2)^2 = (x^2 - y^2)z^2\). (Hint: Show that \((0 : 0 : 1)\) and \((\pm i : 1 : 0)\) are singular points of \(C\). Can it have more singularities? Consider the 1-parameter family of conics from Exercise 2.20.)

EXERCISE 2.23. Find the genus of the curve \(y^2z = x^3 + x^2z\). (Hint: You may apply the same method we used for hyperelliptic curves or you may use a rational parametrization of the curve.)

EXERCISE 2.24. Find the number of tangents to a smooth degree \(n\) curve \(C\) from a generic point \(O\) in \(\mathbb{P}^2\). (Hint: Choose \(O\) so it does not lie on lines that are tangent to \(C\) more than once. Then choose a general enough line \(L\) and consider the projection \(\pi\) of \(C\) onto \(L\) along all the lines through \(O\). Apply the Riemann–Hurwitz formula.)

EXERCISE 2.25. Find the number of tangents to a smooth degree \(n\) curve \(C\) from a generic point \(O\) on \(C\). (Hint: Use Exercise 2.24. What happens to the tangents when \(O\) approaches \(C\)?)
CHAPTER 3

The Riemann–Roch Theorem

3.1. Functions and Local Rings

In this section we will look at some algebraic properties of curves. In particular we will define the ring of functions on a curve and the local ring of a point. This will help us to define discrete valuation, an algebraic analog of the notion of the order of a zero/pole of a function of one complex variable.

3.1.1. Regular and Rational Functions. Let $C$ be a plane affine irreducible curve over $F$ (not necessarily algebraically closed) with equation $f(x, y) = 0$. We would like to define (polynomial) functions on $C$. The idea is that if $g(x, y) = h(x, y) + q(x, y)f(x, y)$ then $g$ and $h$ take the same values at every point of $C$. In other words, we need to consider polynomial function modulo the defining polynomial $f$.

**Definition 3.1.** The quotient ring $F[C] = F[x, y]/\langle f \rangle$ is called the coordinate ring of $C$. Its elements are called regular functions on $C$.

**Example 3.2.**
(a) Let $C$ be the affine line $y = 0$. Then $F[C] = F[x, y]/\langle y \rangle \cong F[x]$, where the last map is $g(x, y) \mapsto g(x, 0)$. This shows that the regular function on $C$ can be identified with the polynomials in $x$.
(b) Let $C$ be the conic with equation $y = x^2$. Then $F[C] = F[x, y]/\langle y - x^2 \rangle \cong F[x]$. Again, the regular functions on $C$ are polynomials in $x$, but the identification is different: $g(x, y) \mapsto g(x, x^2)$.
(c) Let $C$ be a smooth cubic with equation $y^2 = x^3 - x$. Then in the coordinate ring $F[C]$ we have $y^2 = x^3 - x$, and so we can identify $F[C]$ with the set
\[ \{ a_0y + a(x) \mid a_0 \in F, a(x) \in F[x] \} \]

If $f$ is irreducible then the ideal $\langle f \rangle$ is prime. Therefore, if $C$ is irreducible then $F[C]$ is an integral domain. Recall the definition of the field of fractions of an integral domain $R$:

\[ \text{Frac}(R) = \left\{ \frac{a}{b} \mid a, b \in R \right\} , \quad \text{where} \quad \frac{a}{b} = \frac{c}{d} \iff ad = bc \text{ in } R. \]

For example, $\text{Frac}(\mathbb{Z}) = \mathbb{Q}$ and $\text{Frac}(F[x]) = F(x)$, the field of rational functions in $x$. You should check that the operations $\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$ and $\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$ are well-defined and $\text{Frac}(R)$ is indeed a field (the zero and the unit elements are $0 = \frac{0}{1}$ and $1 = \frac{1}{1}$).
3. The Riemann–Roch Theorem

**Definition 3.3.** The function field (a.k.a. the field of rational functions) of $C$ is $\mathbb{F}(C) = \text{Frac}(\mathbb{F}[C])$.

**Example 3.4.**
(a) Let $C$ be the line $y = 0$. Then $\mathbb{F}(C) = \text{Frac}(\mathbb{F}[x]) = \mathbb{F}(x)$.
(b) Let $C$ be the conic $y - x^2 = 0$ and consider the function $y/x \in \mathbb{F}(C)$. What is its value at $(0,0)$? Recall that $y = x^2$ in $\mathbb{F}[C]$, so
$$\frac{y}{x} = \frac{x^2}{x} = x \text{ in } \mathbb{F}(C).$$
Hence we see that this function takes value 0 at $(0,0)$.
(c) Let $C$ be the cubic $y^2 - x^3 + x = 0$. Again, to see the value of $y/x \in \mathbb{F}(C)$ we use that $y^2 = x^3 - x$ in $\mathbb{F}[C]$:
$$\frac{y}{x} = \frac{y^2}{xy} = \frac{x^3 - x}{xy} = \frac{x^2 - 1}{y}.$$
This time we found a representation of $y/x$ where the denominator takes value zero, but the numerator is non-zero at $(0,0)$. In this case we say that the function $y/x$ has a pole at $(0,0)$. Similarly, $x/y = y/(x^2 - 1)$ takes value 0 at $(0,0)$.

In Exercise 3.1 you will show that for any $g \in \mathbb{F}(C)$ its value $g(p) = a(p)/b(p)$ is independent of how you represent $g = a/b$ for $a, b \in \mathbb{F}[C]$ with $b(p) \neq 0$. In other words, if $g = c/d$ for some other $c, d \in \mathbb{F}[C]$ with $d(p) \neq 0$ then $a(p)/b(p) = c(p)/d(p)$ and so the value $g(p)$ is well-defined.

**Definition 3.5.** A rational function $g \in \mathbb{F}(C)$ is called regular at $p \in C$ if there exists a representation $g = a/b$ with $a, b \in \mathbb{F}[C]$ and $b(p) \neq 0$.

This brings us to an important notion of a local ring of a point.

### 3.1.2. The Local Ring of a point.

**Definition 3.6.** The set
$$\mathcal{O}_p = \{ g \in \mathbb{F}(C) \mid g \text{ is regular at } p \}$$
is called the local ring of $p$ on $C$.

It turns out that algebraic properties of $\mathcal{O}_p$ reflect geometric properties of $C$ “near” the point $p$. The following lemma, which you should prove yourself, justifies the word “ring” in the definition of $\mathcal{O}_p$.

**Lemma 3.7.** $\mathcal{O}_p$ is a subring of $\mathbb{F}(C)$, containing $\mathbb{F}[C]$.

Now consider the subset $\mathfrak{m}_p \subset \mathcal{O}_p$ consisting of those $g$ that take value 0 at $p$:
$$\mathfrak{m}_p = \{ g \in \mathcal{O}_p \mid g(p) = 0 \}.$$

It is easy to check that $\mathfrak{m}_p$ is an ideal of $\mathcal{O}_p$. Moreover, we will see that $\mathfrak{m}_p$ is a maximal ideal and it is the only maximal ideal of $\mathcal{O}_p$. In ring theory rings with this property are called local rings.

**Definition 3.8.** A ring $\mathcal{O}$ is called local if $\mathcal{O}$ has a unique maximal ideal $\mathfrak{m}$.

Here is an equivalent definition.
3.1. FUNCTIONS AND LOCAL RINGS

Proposition 3.9. \( O \) is a local ring if and only if \( O \setminus O^* \) is an ideal (and so \( \mathfrak{m} = O \setminus O^* \)). Here \( O^* \) is the set of units of \( O \).

Proof. (\( \Rightarrow \)) Assume \( O \) is a local ring with maximal ideal \( \mathfrak{m} \). First, \( \mathfrak{m} \subseteq O \setminus O^* \) since if an ideal contains a unit then it must contain all elements of the ring, i.e. it is the improper ideal. Second, any \( a \in O \setminus O^* \) defines a proper ideal \( (a) \). Since \( \mathfrak{m} \) is the only maximal ideal of \( O \), we have \( (a) \subset \mathfrak{m} \). But this means that \( a \in \mathfrak{m} \), and we see that \( O \setminus O^* \subseteq \mathfrak{m} \). Therefore, \( \mathfrak{m} = O \setminus O^* \).

(\( \Leftarrow \)) Suppose \( O \setminus O^* \) is an ideal. The same argument as before shows that it must be maximal. Also we saw that any proper ideal must lie in \( O \setminus O^* \). Therefore, any maximal ideal must coincide with \( O \setminus O^* \), i.e. it is a unique maximal ideal.

Example 3.10.

(a) \( \mathbb{Z} \) is not a local ring. Indeed, \( \mathbb{Z} \setminus \{\pm 1\} \) is not an ideal (it is not closed under addition). Also every prime defines a maximal ideal in \( \mathbb{Z} \), so there are infinitely many maximal ideals in \( \mathbb{Z} \).

(b) \( \mathbb{F} \) is a local ring. Its only maximal ideal is the zero ideal (in fact, the only proper ideal). Also \( \{0\} = F \setminus F^* \).

(c) The ring of formal power series \( \mathbb{F}[[x]] = \left\{ g(x) = \sum_{n \geq 0} a_n x^n \mid a_n \in \mathbb{F} \right\} \) is a local ring. Indeed, let \( \mathfrak{m} \) consist of power series \( g(x) \) with \( g(0) = 0 \), i.e. \( a_0 = 0 \). Its complement in \( \mathbb{F}[[x]] \) is the set of power series with \( a_0 \neq 0 \). Such power series are invertible by the following analog of the geometric series formula. Let \( g(x) = a_0(1 - h(x)) \), where \( h \in \mathbb{F}[[x]] \) with \( h(0) = 0 \). Then

\[
\frac{1}{g(x)} = \frac{1}{a_0(1 - h(x))} = \frac{1}{a_0} (1 + h(x) + h(x)^2 + h(x)^3 + \ldots).
\]

Note that every \( x^k \) appears only in the finite number of terms in the infinite sum (all monomials of \( h(x)^m \) for \( m > k \) have degree at least \( k + 1 \) since \( h(0) = 0 \)), so the infinite sum of powers of \( h(x) \) defines a formal power series in \( x \).

Now we are ready to prove that \( O_p \) is a local ring with \( \mathfrak{m}_p \) being the only maximal ideal.

Proposition 3.11. \( O_p \) is a local ring with the maximal ideal \( \mathfrak{m}_p \).

Proof. By definition \( O_p \setminus \mathfrak{m}_p \) consists of \( g = a/b \) with \( a(p) \neq 0 \) and \( b(p) \neq 0 \). Clearly, every such \( g \) has the inverse \( 1/g = b/a \) in \( O_p \). Therefore, \( \mathfrak{m}_p = O_p \setminus O^*_p \). It remains to apply Proposition 3.9. \( \square \)

Remember that the quotient of a ring by a maximal ideal is a field. In the case of the local ring \( O_p \), this field is isomorphic to the base field \( \mathbb{F} \).

Proposition 3.12. Let \( O_p \) be the local ring of a point \( p \in C \). Then \( O_p/\mathfrak{m}_p \cong \mathbb{F} \).

Proof. This follows from the 1st Isomorphism Theorem. Indeed, the map \( \phi : O_p \to \mathbb{F} \), \( g \mapsto g(p) \) is a ring homomorphism with kernel \( \mathfrak{m}_p \). Also \( \phi \) is clearly onto since \( O_p \) contains the constants \( \mathbb{F} \). \( \square \)
Later we will prove that if \( p \) is a non-singular point of \( C \) then the maximal ideal \( m_p \) is, in fact, principal. At the same time we will show that when \( p \) is non-singular, every rational function \( q \in \mathbb{F}(C) \) either lies in \( \mathcal{O}_p \) or its inverse \( 1/g \) lies in \( \mathcal{O}_p \). Right now we will see that this may not be true when \( p \) is a singular point (in fact, never true!).

**Example 3.13.** Let \( C \) be the cuspidal cubic \( y^2 = x^3 \) and let \( p = (0,0) \) be its singular point. Then the function \( y/x \) is not regular at \( p \) and neither is \( x/y \). Indeed, suppose there exists a representation \( y/x = a/b \) with \( a, b \in \mathbb{F}[C] \) and \( b(p) \neq 0 \). Then \( yb = xa \) in \( \mathbb{F}[C] \), which in \( \mathbb{F}[x,y] \) means that

\[
yb = xa + q(y^2 - x^3)
\]

for some polynomials \( a, b, q \in \mathbb{F}[x,y] \). Now the fact that \( b(p) \neq 0 \) implies that \( yb \) has a non-zero \( y \)-term (why?) and, hence, so does \( xa + q(y^2 - x^3) \). But every term in \( xa + q(y^2 - x^3) \) is divisible by \( x \) or is divisible by \( y^2 \), a contradiction. Similarly, you can show that \( x/y \) is not regular at \( p \) (see Exercise 3.2).

In preparation for the proof of the above statements we will prove the following two lemmas.

**Lemma 3.14.** Let \( C \) be a curve defined by \( f(x,y) = 0 \) and \( p = (p_1,p_2) \in C \). Then \( \frac{\partial f}{\partial y}(p) \neq 0 \) if and only if \( \frac{y-p_2}{x-p_1} \in \mathcal{O}_p \).

**Proof.** We may assume that \( p = (0,0) \). The general case is obtained by translating the origin of the coordinate system to \( p = (p_1,p_2) \).

(\(\Rightarrow \)) We will show that \( y/x \in \mathcal{O}_p \). We have

\[
f(x,y) = a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 + \ldots
\]

\[
= a_{10}x + a_{01}y + x^2f_1(x) + xyf_2(x,y) + y^2f_3(y).
\]

By assumption \( \frac{\partial f}{\partial y}(0,0) = a_{01} \neq 0 \). Therefore, in \( \mathbb{F}[C] \) we can write:

\[
y(-a_{01} - xf_2(x,y) - yf_3(y)) = x(a_{10} + xf_1(x)),
\]

which is equivalent to

\[
\frac{y}{x} = \frac{a_{10} + xf_1(x)}{-a_{01} - xf_2(x,y) - yf_3(y)}.
\]

Note that the fraction in the right hand side is defined at \( (0,0) \) since \( a_{01} \neq 0 \). In other words, \( y/x \) is regular at \( p = (0,0) \).

(\(\Leftarrow \)) Suppose \( y/x \) is regular at \( p = (0,0) \). Then \( yb = xa \) for some \( a, b \) in \( \mathbb{F}[C] \) and \( b(p) \neq 0 \). In other words, there exist polynomials \( a, b, q \) in \( \mathbb{F}[x,y] \) such that

\[
yb = xa + qf, \quad b(0,0) \neq 0.
\]

But the latter means that \( b \) has a non-zero constant term and, hence, \( yb \) has a non-zero \( y \)-term. This implies that \( xa + qf \) has a non-zero \( y \)-term, which, of course, can only appear in \( qf \). Therefore,

\[
0 \neq \frac{\partial (qf)}{\partial y}(0,0) = \frac{\partial q}{\partial y}(0,0) \cdot f(0,0) + \frac{\partial f}{\partial y}(0,0) \cdot q(0,0).
\]

Since \( f(0,0) = 0 \) (the point \( p = (0,0) \) lies in \( C \)), we see that \( \frac{\partial f}{\partial y}(0,0) \neq 0 \), as required. \( \square \)
3.1. Functions and Local Rings

Corollary 3.15. If \( p = (p_1, p_2) \in C \) is singular then neither \( \frac{x-p_1}{y-p_2} \) nor \( \frac{y-p_2}{x-p_1} \) is regular at \( p \).

Lemma 3.16. Let \( p = (p_1, p_2) \in C \) and \( m_p \) be the maximal ideal of \( O_p \). Then \( m_p = \langle x-p_1, y-p_2 \rangle \).

Proof. As before we may assume that \( p = (0,0) \). We need to show that \( m_p = \langle x, y \rangle \). Let \( g \in m_p \). Then \( g = a/b \) for some \( a, b \in \mathbb{F}[C] \) and \( a(0) = 0, b(p) \neq 0 \). Let \( a(x, y) \in \mathbb{F}[x, y] \) be a representative of \( a \in \mathbb{F}[C] \). Then \( a(0,0) = 0 \), and so we can write \( a(x, y) = xa_1(x, y) + ya_2(x, y) \) for some \( a_i \in \mathbb{F}[x, y] \). This implies that \( a = xa_1 + ya_2 \in \mathbb{F}[C] \). Now we have

\[
g = \frac{xa_1 + ya_2}{b} = \frac{a_1}{b} + \frac{a_2}{b},
\]

where both \( a_1/b \) and \( a_2/b \) lie in \( O_p \) as \( b(p) \neq 0 \). In other words, \( g \in \langle x, y \rangle \).

It remains to recall that \( m_p \) is a maximal ideal and \( \langle x, y \rangle \) is a proper ideal, so \( m_p \subseteq \langle x, y \rangle \) implies \( m_p = \langle x, y \rangle \). \( \square \)

We are ready to prove the first statement we made before Example 3.13.

Proposition 3.17. Let \( C \) be an affine curve over \( \mathbb{F} \). If \( p \in C \) is a non-singular point then \( m_p \) is a principal ideal.

Proof. If \( p \) is non-singular then either \( \frac{\partial f}{\partial x}(p) \neq 0 \) or \( \frac{\partial f}{\partial y}(p) \neq 0 \). Without loss of generality we will assume the latter. Then by Lemma 3.14 \( g = \frac{y-p_2}{x-p_1} \in O_p \). Therefore, by Lemma 3.16,

\[
m_p = \langle x-p_1, y-p_2 \rangle = \langle x-p_1, (x-p_1)g \rangle = \langle x-p_1 \rangle,
\]

i.e. \( m_p \) is principal. \( \square \)

Remark 3.18. As it follows from the above proof, if \( \frac{\partial f}{\partial y}(p) \neq 0 \) then \( t = x-p_1 \) generates \( m_p \) and if \( \frac{\partial f}{\partial x}(p) \neq 0 \) then \( t = y-p_2 \) generates \( m_p \). Of course, we should keep in mind that \( x-p_1 \) and \( y-p_2 \) denote elements of \( O_p \), not simply polynomials.

3.1.3. Local Parameters.

Definition 3.19. We say that \( t \in O_p \) is a local parameter if for any \( g \in O_p \) we can write \( g = ut^m \) for some unit \( u \in O_p^* \) and some non-negative integer \( m \).

Note that if \( t \) is a local parameter of \( O_p \) then \( t \) generates the maximal ideal \( m_p \). Indeed, every \( g \in m_p = O \setminus O^* \) can be written as \( ut^m \) for some \( m \geq 1 \), i.e. \( g \in \langle t \rangle \).

In the next theorem we show that local parameters exist in the local ring of a non-singular point.

Theorem 3.20. Let \( C \) be an affine curve and let \( p \in C \) be non-singular. Then there exists a local parameter \( t \in O_p \).

Proof. First, we may assume that \( p = (0,0) \) and \( \frac{\partial f}{\partial y}(p) \neq 0 \). By Remark 3.18 \( m_p = \langle x \rangle \). Now consider a chain of ideals:

\[
O_p \supseteq m_p \supseteq \langle x \rangle \supseteq \langle x^2 \rangle \supseteq \langle x^3 \rangle \supseteq \ldots
\]

Krull’s intersection theorem says that the intersection \( \cap_{i \geq 1} \langle x^i \rangle \) is zero. Therefore, for any non-zero \( g \in O_p \) there exists \( m \) such that \( g \in \langle x^m \rangle \), but \( g \not\in \langle x^{m+1} \rangle \). Then \( g = xu^m \) for some \( u \in O_p \). Since \( u \not\in \langle x \rangle = m_p \) (otherwise \( g \in \langle x^{m+1} \rangle \)) we see that \( u \in O_p \setminus m_p = O_p^* \), i.e. \( u \) is a unit. \( \square \)
Remark 3.21. We see from the proof and Remark 3.18 that if $\frac{\partial f}{\partial x}(p) \neq 0$ then we can choose $t = x - p_1$ to be a local parameter. Similarly, if $\frac{\partial f}{\partial x}(p) \neq 0$ then $t = y - p_2$ is a local parameter. Also, we see that any two local parameters $t, t'$ differ by a unit multiple: $t' = ut$ for some $u \in \mathcal{O}_p$, since each of them generates $m_p$.

Theorem 3.22. Let $C$ be an affine irreducible curve and let $p \in C$. Then $p$ is non-singular if and only if for any $g \in \mathbb{F}(C)$ either $g \in \mathcal{O}_p$, or $1/g \in \mathcal{O}_p$.

Proof. ($\Leftarrow$) Follows from Corollary 3.15.

($\Rightarrow$) Assume $p \in C$ is non-singular. By Theorem 3.20 there exists a local parameter $t \in \mathcal{O}_p$. Now for any $g \in \mathbb{F}(C)$ we have $g = a/b$ for some $a, b \in \mathcal{O}_p$, so $a = u_1 t^{m_1}$ and $b = u_2 t^{m_2}$, where $u_i \in \mathcal{O}_p$ and $m_i \in \mathbb{Z}_{\geq 0}$. Then $g = u_1 u_2^{-1} t^{m_1 - m_2}$. If $m_1 \geq m_2$ then $g \in \mathcal{O}_p$, if $m_1 < m_2$ then $1/g \in \mathcal{O}_p$. 

3.1.4. Discrete Valuation.

Definition 3.23. Let $\mathbb{L}$ be a field. A surjective map $v : \mathbb{L}^* \to \mathbb{Z}$ is called a discrete valuation if

(a) $v(fg) = v(f) + v(g),$
(b) $v(f + g) \geq \min(v(f), v(g)).$

We will use a convention $v(0) = \infty$.

It turns out that with every non-singular point $p \in C$ we get a discrete valuation on the field of rational functions $\mathbb{F}(C)$.

Theorem 3.24. Let $C$ be an affine irreducible curve and let $p \in C$ be non-singular. Then the map

$$v_p : \mathbb{F}(C)^* \to \mathbb{Z}, \quad g \mapsto m,$$

where $g = ut^m$ for $u \in \mathcal{O}_p$, $m \in \mathbb{Z}$, is a discrete valuation.

Proof. By Theorem 3.20 and Theorem 3.22, there exists $t \in \mathcal{O}_p$ such that every $g \in \mathbb{F}(C)$ can be written as $g = ut^m$ for some $u \in \mathcal{O}_p^*$ and $m \in \mathbb{Z}$. You should check that the map $v_p$ does not depend on the choice of $t$ (use the fact that any two local parameters differ by a unit multiple). Now suppose $g_1 = u_1 t^{m_1}$ and $g_2 = u_2 t^{m_2}$ and assume $m_1 \leq m_2$. Then

$$g_1 g_2 = u_1 u_2 t^{m_1 + m_2}, \quad \text{so} \quad v_p(g_1 g_2) = v_p(g_1) + v_p(g_2).$$

Also we have $g_1 + g_2 = (u_1 + u_2 t^{m_1 - m_2 - 1}) t^{m_1}$. Note that $u_1 + u_2 t^{m_1 - m_2 - 1}$ lies in $\mathcal{O}_p$, so $v_p(u_1 + u_2 t^{m_1 - m_2}) \geq 0$. Therefore,

$$v_p(g_1 + g_2) = v_p(u_1 + u_2 t^{m_1 - m_2}) + v_p(t^{m_1}) \geq m_1 = \min(v_p(g_1), v_p(g_2)).$$

Finally, $v_p$ is surjective as $v_p(t^m) = m$ for any $m \in \mathbb{Z}$ and $t^m \in \mathbb{F}(C)$. 

Now we will consider the case when $C$ is an irreducible projective curve. We can translate all the above definition as follows. Let $C$ be defined by a homogeneous polynomial $F(x, y, z)$.

Definition 3.25. The quotient ring $\mathbb{F}[C] = \mathbb{F}[x, y, z]/(F)$ is called the homogeneous coordinate ring of $C$. The field of rational functions $\mathbb{F}(C)$ is

$$\mathbb{F}(C) = \left\{ \frac{G}{H} \mid G, H \in \mathbb{F}[x, y, z]_d, d \in \mathbb{Z}_{\geq 0} \right\},$$
where \( G/H = G'/H' \) if and only if \( GH' = HG' \) in \( \mathbb{F}[C] \), i.e. \( GH' - HG' \in \langle F \rangle \). Here \( \mathbb{F}[x, y, z]_d \) is the subspace of homogeneous polynomials of degree \( d \).

The reason we need the degree of \( G \) and \( H \) to be the same is so the value of \( g = G/H \) at a point \( (x : y : z) \in \mathbb{P}^2 \) is well-defined. Indeed, if \( p = (x : y : z) = (\lambda x : \lambda y : \lambda z) \) then

\[
g(p) = \frac{G(\lambda x, \lambda y, \lambda z)}{H(\lambda x, \lambda y, \lambda z)} = \frac{\lambda^d G(x, y, z)}{\lambda^d H(x, y, z)} = \frac{G(x, y, z)}{H(x, y, z)}.
\]

Just as in the affine case you should check that the value \( g(p) \) does not depend on the representation \( g = G/H \).

**Definition 3.26.** Let \( p \in C \). We say that \( g \in \mathbb{F}(C) \) is regular at \( p \) if we can write \( g = G/H \) such that \( H(p) \neq 0 \). They form the local ring \( \mathcal{O}_p \subset \mathbb{F}(C) \) at the point \( p \).

As before \( \mathcal{O}_p \) has a unique maximal ideal \( \mathfrak{m}_p \). If \( p \) is a non-singular point of \( C \) and it lies in the affine part \( z = 1 \) then we can choose either \( t = \frac{x-p_1z}{z} \) or \( t = \frac{y-p_2z}{z} \) to be a local parameter (a generator for \( \mathfrak{m}_p \)). Therefore we can define the discrete valuation \( v_p : \mathbb{F}(C) \to \mathbb{Z} \) in the same way we did it in Theorem 3.24.

**Example 3.27.** Let \( C \) be the conic \( F(x, y, z) = yz - x^2 = 0 \) in \( \mathbb{P}^2 \) and consider the rational function \( g = y/x \) in \( \mathbb{F}(C) \).

(a) Let \( p = (0 : 0 : 1) \). What is \( v_p(g) \)? We have

\[
g = \frac{y}{x} = \frac{xy}{x^2} = \frac{yz}{y} = \frac{x}{z},
\]

which shows that \( g \) is regular at \( p \). Now \( p \) lies in the affine part \( z = 1 \), where \( C \) is defined by \( f(x, y) = F(x, y, 1) = y - x^2 \). Note that \( \frac{\partial f}{\partial y} = 1 \neq 0 \), so \( x/z \in \mathbb{F}(C) \) is a local parameter at \( p \). Therefore \( v_p(g) = 1 \).

(b) Now let us consider the infinite point \( p = (0 : 1 : 0) \) on \( C \). It lies in the affine part \( y = 1 \) where \( C \) is defined by \( f(x, z) = z - x^2 \). Similarly to the previous case \( \frac{\partial f}{\partial z} = 1 \neq 0 \), so \( x/y \) is a local parameter at \( p \). Since \( g = (x/y)^{-1} \) we have \( v_p(g) = -1 \). In this case we will say that \( g \) has a pole of order \( 1 \) at \( p \).

**Example 3.28.** Let \( C \) be the conic \( F(x, y, z) = yz - x^2 = 0 \) in \( \mathbb{P}^2 \) as before, and consider the rational function \( g = \frac{x-z}{x+z} \) in \( \mathbb{F}(C) \).

(a) Let \( p_1 = (1 : 0 : 1) \) and \( p_2 = (-1 : 0 : 1) \). We can choose \( \frac{x-z}{x+z} \) to be a local parameter at \( p_1 \). Then

\[
g = \frac{x-z}{x+z} = \frac{z}{x+z} \left( \frac{x-z}{z} \right).
\]

Note that the first factor in the right hand side is a unit in \( \mathcal{O}_{p_2} \) (it takes a non-zero value at \( p_1 \)), hence, by definition \( v_{p_1}(g) = 1 \).

Similarly, \( \frac{x+z}{z} \) is a local parameter at \( p_2 \) and we have

\[
g = \frac{x-z}{x+z} = \frac{x-z}{z} \left( \frac{x+z}{z} \right)^{-1},
\]

so \( v_{p_2}(g) = -1 \).

(c) Let \( p = (0 : 1 : 0) \). Since \( g \) is regular and non-zero at \( p \) it lies in \( \mathcal{O}_p \setminus \mathfrak{m}_p \), i.e. is a unit. Therefore \( v_p(g) = 0 \).
3. THE RIEMANN–ROCH THEOREM

**Definition 3.29.** Let $C$ be an irreducible curve and $p$ a non-singular point on $C$. For $g \in \mathbb{F}(C)$ we call $\nu_p(g)$ the order of $g$ at $p$. If $\nu_p(g) = k > 0$ we say that $g$ has a zero of order $k$ at $p$. If $\nu_p(g) = -k < 0$ we say that $g$ has a pole of order $k$ at $p$.

The following theorem shows the difference between the affine and the projective case.

**Theorem 3.30.** Let $\mathbb{K}$ be an algebraically closed field.

1. If $C$ is an affine curve over $\mathbb{K}$ then the set of rational functions regular at every point of $C$ coincides with $\mathbb{K}[C]$.
2. If $C$ is a projective curve over $\mathbb{K}$ then the set of rational functions regular at every point of $C$ consists of constants only, i.e. equals $\mathbb{K}$.

Although we will not prove this important theorem it should remind you of Liouville’s theorem from complex analysis which says that the only bounded entire functions (which are the regular functions on $\mathbb{P}^1_C$) are constants.

### 3.2. Divisors

**3.2.1. The divisor group.** Let $C$ be a smooth irreducible projective curve over $\mathbb{K}$, an algebraically closed field.

**Definition 3.31.** A divisor $D$ on $C$ is a formal sum

$$D = \sum_{p \in C} a_p p,$$

where $a_i \in \mathbb{Z}$ and $a_p = 0$ for all but finitely many $p \in C$.

You can think of this as an abstract generalization of the notion of the intersection of two curves that we studied in Section 2.4. Indeed, $C \cap E$ consists of finitely many points $p \in C$ with some multiplicities (intersection numbers) $a_p = (C \cdot E)_p \in \mathbb{Z}$. Here are a few related definitions.

**Definition 3.32.** The support of a divisor $D = \sum_{p \in C} a_p p$ is the set

$$\text{Supp} \, D = \{ p \in C \mid a_p \neq 0 \}.$$

The degree of $D$ is the integer

$$\deg D = \sum_{p \in C} a_p.$$

**Proposition 3.33.** The set of all divisors on $C$ forms an Abelian group $\text{Div}(C)$. The degree map

$$\deg : \text{Div}(C) \to \mathbb{Z}, \quad D \mapsto \deg D$$

is a surjective group homomorphism.

**Proof.** First, naturally we can define the addition operation on $\text{Div}(C)$: If $D_1 = \sum_{p \in C} a_p p$ and $D_2 = \sum_{p \in C} b_p p$. Then

$$D_1 + D_2 = \sum_{p \in C} (a_p + b_p) p \in \text{Div}(C).$$
The identity element is $O = \sum_{p \in C} a_p p$ with $a_p = 0$ for all $p \in C$ (i.e. the divisor with empty support) and the additive inverse of $D = \sum_{p \in C} a_p p$ is

$$-D = \sum_{p \in C} (-a_p) p.$$ 

You should check all the group axioms yourself.

Second, the degree map is a group homomorphism:

$$\deg(D_1 + D_2) = \sum_{p \in C} (a_p + b_p) = \sum_{p \in C} a_p + \sum_{p \in C} b_p = \deg D_1 + \deg D_2.$$ 

Also it is surjective, since for any $m \in \mathbb{Z}$ the divisor $m p$ (where $p$ is some fixed point of $C$) has degree $m$. \hfill $\square$

**Definition 3.34.** A divisor $D = \sum_{p \in C} a_p p$ is effective if $a_p \geq 0$ for every $p \in C$. We will write $D \geq 0$ to indicate that $D$ is effective.

**Example 3.35.** Let $g \in \mathbb{K}(C)$ be a rational function not identically zero. Define the divisor of $g$

$$(g) = \sum_{p \in C} v_p(g) p,$$

where $v_p(g)$ is the order of $g$ at $p \in C$. We also define the divisor of zeroes of $g$ and the divisor of poles of $g$

$$(g)_0 = \sum_{v_p(g) > 0} v_p(g) p, \quad (g)_\infty = \sum_{v_p(g) < 0} (-v_p(g)) p.$$ 

Note that both $(g)_0$ and $(g)_\infty$ are effective divisors and $(g) = (g)_0 - (g)_\infty$.

**Definition 3.36.** A divisor $D \in \text{Div}(C)$ is called principal if $D = (g)$ for some rational function $g \in \mathbb{K}(C)^*$.  

**Proposition 3.37.** The degree of a principal divisor equals zero.

**Proof.** Let $D = (g)$ where $g = G/H$ for some homogeneous polynomials $G, H \in \mathbb{K}[x, y, z]$ of the same degree $m$. Let $C_G$ and $C_H$ be the projective curves defined by $G$ and $H$, respectively. Then

$$(g)_0 = \sum_{p \in C_G \cap C} (C_G \cdot C)_p p, \quad (g)_\infty = \sum_{p \in C_H \cap C} (C_H \cdot C)_p p.$$ 

By Bézout's theorem $\deg(g)_0 = \sum (C_G \cdot C)_p = mn$ and $\deg(g)_\infty = \sum (C_H \cdot C)_p = mn$, where $n = \deg C$. Therefore $\deg(g) = mn - mn = 0$. \hfill $\square$

Now we are going to define an equivalence relation on the group of divisors.

**Definition 3.38.** Two divisors $D_1, D_2 \in \text{Div}(C)$ are called equivalent if $D_1 - D_2$ is principal. We will write $D_1 \sim D_2$ in this case.

In fact, the set of principal divisors forms a subgroup $P(C)$ of $\text{Div}(C)$ and the equivalence classes of divisors are the elements of the quotient group $\text{Div}(C)/P(C)$.

**Proposition 3.39.** The set of principal divisors $P(C)$ forms a subgroup of $\text{Div}(C)$.
PROOF. First, \( P(C) \) is non-empty since \( \mathbb{K}(C)^\ast \) is non-empty. For example, every non-zero constant \( c \) defines the zero divisor \( (c) = O \) in \( \text{Div}(C) \). Now for any \( D_1, D_2 \in P(C) \) we have

\[
D_1 - D_2 = (g_1) - (g_2) = (g_1/g_2).
\]

Indeed, \( v_p(g_1/g_2) = v_p(g_1) + v_p(1/g_2) = v_p(g_1) - v_p(g_2) \) for any \( p \in C \), by the properties of the valuation. Therefore, \( D_1 - D_2 \in P(C) \), so \( P(C) \) is a subgroup. \( \square \)

**Definition 3.40.** The quotient group \( \text{Div}(C)/P(C) \) is called the class group of \( C \) and is denoted by \( \text{Cl}(C) \). We will also write \([D]\) to denote the class in \( \text{Cl}(C) \) containing \( D \).

Putting all these definitions together we see that \( D_1 \sim D_2 \) if and only if \( D_1 = D_2 + (g) \) for some \( g \in \mathbb{K}(C)^\ast \) if and only if \([D_1] = [D_2]\) in \( \text{Cl}(C) \).

We also remark that the degree homomorphism \( \text{deg}: \text{Div}(C) \to \mathbb{Z} \) induces the degree homomorphism \( \text{deg}: \text{Cl}(C) \to \mathbb{Z} \) defined by \( \text{deg}([D]) = \text{deg} D \). It is well-defined since \( \text{deg}(D + (g)) = \text{deg} D + \text{deg}(g) = \text{deg} D \), by Proposition 3.37.

**3.2.2. The Riemann–Roch Space of a divisor.**

**Definition 3.41.** Let \( D \in \text{Div}(C) \). The Riemann–Roch space of \( D \) is the set

\[
\mathcal{L}(D) = \{ f \in \mathbb{K}(C)^\ast \mid (f) + D \geq 0 \} \cup \{0\}.
\]

In other words, \( \mathcal{L}(D) \) consists of rational functions \( f \) which make the divisor \((f) + D\) effective. More explicitly, if we write \( D = \sum_{p \in C} a_p p \) as a difference of two effective divisors

\[
D = \sum_{a_p > 0} a_p p - \sum_{a_p < 0} (-a_p) p = D_+ - D_-, \quad D_+ \geq 0, D_- \geq 0,
\]

then \( f \in \mathbb{K}(C)^\ast \) if and only if it has a zero at every \( p \in \text{Supp} D_- \) of order at least \( a_p \) and a pole at every \( p \in \text{Supp} D_+ \) of order at most \( a_p \).

**Example 3.42.** Let \( C \) be the projective line \( y = 0 \) in \( \mathbb{P}^2 \).

(a) Consider the divisor \( D = 3 \cdot \infty - 2 \cdot 1 \), where \( 1 = (1 : 0 : 1) \) and \( \infty = (1 : 0 : 0) \). Then \( D = D_+ - D_- \) where \( D_+ = 3 \cdot \infty \) and \( D_- = 2 \cdot 1 \). A rational function \( f \) lies in \( \mathcal{L}(D) \) if and only if it has a zero at \( 1 \) of order at least 2 and a pole at \( \infty \) of order at most 3. Note that \( f \) may have zeroes at other points, but no more poles. We can write such \( f \) explicitly:

\[
f(x, z) = \frac{G(x, z)(x-z)^{k_1}}{z^{k_2}}, \quad k_1 \geq 2, \ k_2 \leq 3,
\]

for some homogeneous \( G \in \mathbb{K}[x, z] \) of degree \( k_2 - k_1 \). The above inequalities imply that \( 0 \leq \text{deg} G \leq 1 \) and, hence, we can write

\[
f(x, y) = \frac{(a_0 z + a_1 x)(x-z)^2}{z^3}, \quad a_0, a_1 \in \mathbb{K}.
\]

In particular, we see that \( \mathcal{L}(D) \) is a 2-dimensional vector space over \( \mathbb{K} \)

\[
\mathcal{L}(D) = \text{span}_\mathbb{K} \left\{ \frac{(x-z)^2}{z^2}, \ \frac{x(x-z)^2}{z^3} \right\}.
\]
Note that we can also use rational functions in a local parameter to describe \( \mathcal{L}(D) \). If we put \( t = x/z \), a local parameter at the origin, then we can write

\[
\mathcal{L}(D) = \text{span}\{ (t-1)^2, t(t-1)^2 \}.
\]

(b) Now let \( D = m \cdot \infty, m \in \mathbb{Z} \). Then \( f \in \mathcal{L}(D) \) if and only if it has the only pole at \( \infty \) of order at most \( m \). Thus we can write

\[
\mathcal{L}(D) = \left\{ \frac{F(x,z)}{z^k} \mid \deg F = k, k \leq m \right\}.
\]

Alternatively, using \( t = x/z \) to dehomogenize the above polynomials, we can write

\[
\mathcal{L}(D) = \{ f(t) \in \mathbb{K}[t] \mid \deg f \leq m \},
\]

so \( \mathcal{L}(D) \) is a vector space over \( \mathbb{K} \) of dimension \( m + 1 \).

In general, finding the dimension of \( \mathcal{L}(D) \) is tricky. Later we will introduce the Riemann–Roch theorem that can help to compute \( \dim \mathcal{L}(D) \). Right now we will prove that \( \dim \mathcal{L}(D) \) is always finite by estimating it from above.

**Theorem 3.43.** Let \( D \) be a divisor on \( C \). Then \( \mathcal{L}(D) \) is a finite dimensional vector space over \( \mathbb{K} \) of dimension at most \( \deg D + 1 \).

**Proof.** First, let us check that \( \mathcal{L}(D) \) is indeed a vector space. By definition \( 0 \in \mathcal{L}(D) \). Next, for any \( f \in \mathcal{L}(D) \) and any \( c \in \mathbb{K}^* \) we have \( (cf) = (f) \), so \( cf \in \mathcal{L}(D) \). Now let \( f_1, f_2 \in \mathcal{L}(D) \) and so \( (f_1) + D \geq 0, (f_2) + D \geq 0 \). More explicitly, let \( D = \sum_{p \in C} a_p p \), then \( v_p(f_1) \geq -a_p \) for \( i = 1, 2 \) and for every \( p \in C \).

On the other hand, by the properties of a discrete valuation

\[
v_p(f_1 + f_2) \geq \min(v_p(f_1), v_p(f_2)) \geq -a_p, \quad \forall p \in C,
\]

and so \( (f_1 + f_2) + D \geq 0 \). Therefore, \( f_1 + f_2 \in \mathcal{L}(D) \).

To prove the estimate on \( \dim \mathcal{L}(D) \) we first show that it is enough to assume that \( D \) is effective and then use induction on \( \deg D \geq 0 \). Indeed, if we write \( D = D_+ - D_- \) with effective \( D_+ \), \( D_- \) then it is easy to see that \( \mathcal{L}(D) \subseteq \mathcal{L}(D_+) \) (check that!) and, hence, an upper bound for \( \dim \mathcal{L}(D_+) \) serves as an upper bound for \( \dim \mathcal{L}(D) \) as well.

Now assume \( D \) is effective. The base of induction is simple: If \( D \geq 0 \) and \( \deg D = 0 \) then \( D = 0 \). In this case

\[
\mathcal{L}(0) = \{ f \in \mathbb{K}(C)^* \mid (f) \geq 0 \} \cup \{ 0 \} = \{ f \in \mathbb{K}(C)^* \mid v_p(f) \geq 0, \forall p \in C \} \cup \{ 0 \},
\]

i.e., \( \mathcal{L}(0) \) consists of rational functions which are regular everywhere on \( C \). By Theorem 3.30, they are only constants, so \( \mathcal{L}(0) = \mathbb{K} \). Therefore, \( \dim \mathcal{L}(0) = 1 = \deg D + 1 \).

Suppose \( D \geq 0 \) and \( \deg D > 0 \). Let \( p \in \text{Supp}(D) \), so \( a_p > 0 \), and let \( t_p \) be a local parameter at \( p \in C \). For any \( f \in \mathcal{L}(D) \) we have \( v_p(f) + a_p \geq 0 \), hence, \( v_p(t_p^a f) \geq 0 \). In other words, \( t_p^a f \) is regular at \( p \) and so its value at \( p \) is defined. Consider the map

\[
\phi : \mathcal{L}(D) \to \mathbb{K}, \quad f \mapsto (t_p^a f)(p).
\]

This is clearly a linear map with the kernel

\[
\ker \phi = \{ f \in \mathcal{L}(D) \mid v_p(t_p^a f) > 0 \} = \{ f \in \mathcal{L}(D) \mid v_p(f) + (a_p - 1) \geq 0 \} \subseteq \mathcal{L}(D - p).
\]
Since $\deg(D - p) < \deg D$ we may apply the inductive hypothesis:
\[
\dim \mathcal{L}(D - p) \leq \deg(D - p) + 1 = \deg D.
\]

Now from linear algebra
\[
\dim \mathcal{L}(D) = \dim \text{Ker} \phi + \dim \text{Im} \phi \leq \dim \mathcal{L}(D - p) + \dim K \leq \deg D + 1.
\]

\[\square\]

From now on we will be denoting the dimension of $\mathcal{L}(D)$ by $\ell(D)$. Next we will show that $\ell(D)$ is invariant of the class of $D$ in $\text{Cl}(C)$.

**Proposition 3.44.** $\ell(D)$ is independent of the choice of $D$ in $[D] \in \text{Cl}(C) = \text{Div}(C)/\mathbb{P}(C)$.

**Proof.** Suppose $D_1 \sim D_2$ in $\text{Cl}(C)$, i.e. $D_1 = (f_0) + D_2$ for some $f_0 \in K(C)$. Consider the map
\[
\phi : \mathcal{L}(D_1) \to \mathcal{L}(D_2), \quad f \mapsto f_0 f.
\]
This is well-defined since $f \in \mathcal{L}(D_1)$ implies $(f) + D_1 \geq 0$, i.e. $(f) + (f_0) + D_2 \geq 0$, and so $(f_0 f) + D_2 \geq 0$, which means $f_0 f \in \mathcal{L}(D_2)$. Also the map is linear. On the other hand
\[
\phi^{-1} : \mathcal{L}(D_2) \to \mathcal{L}(D_1), \quad g \mapsto f_0^{-1} g
\]
is the inverse of $\phi$, hence, $\phi$ is an isomorphism of vector spaces, $\phi : \mathcal{L}(D_1) \cong \mathcal{L}(D_2)$. Therefore, $\ell(D_1) = \ell(D_2)$. \[\square\]

**Example 3.45.** Once again, let $C$ be the projective line $\mathbb{P}^1$. This time we will use a local parameter $t$ (e.g. if $C$ is given by $y = 0$ then we can take $t = x/z$ to be a local parameter at the origin). Since $\mathbb{P}^1 = \mathbb{A}^1 \cup \{\infty\}$, any divisor on $\mathbb{P}^1$ can be written as
\[
D = a_1 \cdot p_1 + \cdots + a_k \cdot p_k + a_0 \cdot \infty
\]
for some $p_i \in \mathbb{A}^1$ and $a_i \in \mathbb{Z}$. Now consider
\[
f(t) = (t - p_1)^{a_1} \cdots (t - p_k)^{a_k} \in K(t).
\]
Then
\[
(f) = (f)_0 - (f)_\infty = a_1 \cdot p_1 + \cdots + a_k \cdot p_k - \left( \sum_{i=1}^{k} a_i \right) \cdot \infty.
\]
This implies that
\[
D - (f) = \left( \sum_{i=0}^{k} a_i \right) \cdot \infty = (\deg D) \cdot \infty.
\]
Therefore, any divisor $D$ on $\mathbb{P}^1$ is equivalent to $(\deg D) \cdot \infty$.

This observation produces the following formula for $\ell(D)$ for any effective divisor on $\mathbb{P}^1$. This is, in fact, an instance of the Riemann–Roch formula which we will discuss in Section 3.4.

**Proposition 3.46.** Let $D$ be an effective divisor on $\mathbb{P}^1$. Then $\ell(D) = \deg D + 1$.

**Proof.** We have seen in Proposition 3.44 that $\ell(D)$ is independent of the choice of $D$ in $[D]$. On the other hand, $D$ is equivalent to $(\deg D) \cdot \infty$. Also, by Example 3.42, $\ell((\deg D) \cdot \infty) = \deg D + 1$. \[\square\]

In particular, we see that the bound on $\ell(D)$ from Theorem 3.43 is attained on $\mathbb{P}^1$. 
3.3. Differential Forms

In this section we will define differential forms on algebraic curves, which is a generalization of the usual differential $dx$ from calculus or complex analysis. The idea is that a differential form may look different in different part of the curve. We will start with an example of a differential form on $\mathbb{P}^1$. You may find this construction familiar if you have studied the residue theory in complex analysis.

**Example 3.47.** Consider the projective line $\mathbb{P}^1 = \mathbb{A}^1 \cup \{\infty\}$ with homogeneous coordinates $(x : z)$. We can choose $t = x/z$ for a local parameter at the origin $(0 : 1)$ Then $u = 1/t = z/x$ is a local parameter at infinity $(1 : 0)$. Now the differential $dt$ is defined in the affine part $U_1 = \{z \neq 0\}$ and the differential $-\frac{1}{t^2}du$ is defined in the affine part $U_2 = \{x \neq 0\}$. Moreover, in the intersection $U_1 \cap U_2$ they coincide:

$$dt = d\left(\frac{1}{u}\right) = -\frac{1}{u^2}du.$$  

In this situation we can say that they define a differential form $\omega$ on $\mathbb{P}^1$.

Let us see how this can be generalized.

### 3.3.1. Differential Forms.

First, we would like to define what $df$ is when $f$ is a regular function on a curve $C$. Our definition is motivated by the familiar notion of the Taylor series. Let $f$ be a rational function in a variable $t$. If it is regular at $p \in \mathbb{P}^1$ then it has a Taylor series expansion

$$f = f(p) + f'(p)t + \frac{f''(p)}{2}t^2 + \ldots$$

The second term $d_pf = f'(p)t$ is a linear function in $t$ which is the value of the differential $df$ at $p$. We can extract it by considering $f - f(p)$ modulo higher order terms, i.e. terms divisible by $t^2$.

Now let $C$ be a smooth curve over $\mathbb{K}$ and $p \in C$. Then there exists a local parameter $t \in \mathbb{K}(C)$ such that $(t) = m_p \subset \mathcal{O}_p$. Now for any $f \in \mathcal{O}_p$ we have

$$f - f(p) \in m_p = (t).$$

Note that $m_p \supset m_p^2 = (t^2)$ and we can consider the quotient space $m_p/m_p^2 = (t)/(t^2)$, called the cotangent space at $p$.

**Definition 3.48.** The differential $d_pf$ at $p$ is the class of $f - f(p)$ in $m_p/m_p^2$, i.e.

$$d_pf = f - f(p) \mod m_p^2.$$ 

This defines a map

$$d_p : \mathcal{O}_p \to m_p/m_p^2, \quad f \mapsto d_pf.$$ 

**Proposition 3.49.** The map $d_p$ is a linear map of $\mathbb{K}$-vector spaces which satisfies the Leibniz rule

$$d_p(f \cdot g) = d_pf \cdot g + f \cdot d_pg.$$ 

**Proof.** First, for any $f, g \in \mathcal{O}_p$ and $c \in \mathbb{K}$ we have

$$d_p(f+g) = f+g - (f(p)+g(p)) \mod m_p^2 = (f-f(p))+(g-g(p)) \mod m_p^2 = d_pf+d_pg,$$

$$d_p(cf) = cf - cf(p) \mod m_p^2 = c(f - f(p)) \mod m_p^2 = c d_pf,$$

so $d_p$ is a linear map. Now

$$d_p(f \cdot g) = f \cdot g - (f(p) \cdot g(p)) \mod m_p^2.$$
On the other hand,
\[ d_p f \cdot g + f \cdot d_p g = (f - f(p))g + f(g - g(p)) \mod \mathfrak{m}_p^2 \]
(3.2)
\[ 2f \cdot g - f(p) \cdot g - f \cdot g(p) \mod \mathfrak{m}_p^2. \]
Subtracting (3.1) from (3.2) and factoring the right hand side we obtain
\[ d_p(f \cdot g) - (d_p f \cdot g + f \cdot d_p g) = (f - f(p))(g - g(p)) \mod \mathfrak{m}_p^2. \]
Since both \( f - f(p) \) and \( g - g(p) \) lie in \( \mathfrak{m}_p \), their product lies in \( \mathfrak{m}_p^2 \). Therefore,
\[ d_p(f \cdot g) - (d_p f \cdot g + f \cdot d_p g) = 0 \mod \mathfrak{m}_p^2, \]
as required.

**Example 3.50.** Let \( t \) be a local parameter at a non-singular point \( p \in C \). By definition, \( d_p t = t - t(p) = t \mod \mathfrak{m}_p^2 \). So we see that \( d_p t \) spans the 1-dimensional cotangent space \( \mathfrak{m}_p/\mathfrak{m}_p^2 = \langle t \rangle/(t^2) \).

Now we are ready to define differential forms in general.

**Definition 3.51.** Let \( C \) be a smooth projective curve and let \( C = U_1 \cup \cdots \cup U_k \), where each \( U_i = C \setminus \{\text{finite set of pts}\} \) is an open subset. Suppose on every \( U_i \) we assign \( f_i dt_i \), where \( f_i \in \mathbb{K}(C) \) and for every \( p \in U_i \) the function \( t_i - t_i(p) \) is a local parameter at \( p \in U_i \). Suppose furthermore that \( f_i dt_i = f_j dt_j \) on every intersection \( U_i \cap U_j \). Then the collection of open sets \( U_i \) together with \( f_i dt_i \) is called a rational differential form \( \omega \) on \( C \). We will write \( \omega = f_i dt_i \) on \( U_i \).

If for every \( t \) the function \( f_t \) is regular on \( U_t \) then the form \( \omega \) is called a regular differential form on \( C \).

We set the following notation:
\[ \Omega(C) \text{ is the } \mathbb{K}\text{-vector space of all rational differential forms on } C, \]
\[ \Omega[C] \text{ is the } \mathbb{K}\text{-vector space of all regular differential forms on } C. \]

**Example 3.52.** Let \( C \) be a projective cubic given by \( X^3 + Y^3 + Z^3 = 0 \). (We use \( (X : Y : Z) \) to denote the homogeneous coordinates in \( \mathbb{P}^2 \).) Consider the open sets
\[ U_1 = \{(X : Y : Z) \in C \mid Y \neq 0, Z \neq 0\}, \]
\[ U_2 = \{(X : Y : Z) \in C \mid X \neq 0, Z \neq 0\}, \]
\[ U_3 = \{(X : Y : Z) \in C \mid X \neq 0, Y \neq 0\}. \]
The union \( U_1 \cup U_2 \) is the affine part of \( C \) with affine coordinates \( x = X/Z \) and \( y = Y/Z \) and the affine equation \( x^3 + y^3 + 1 = 0 \). Furthermore, \( dx/y^2 \) is regular on \( U_1 \) and \( -dy/x^2 \) is regular on \( U_2 \). Moreover, they coincide on the intersection \( U_1 \cap U_2 \). Indeed, on \( U_1 \cap U_2 \) the curve is given by \( x^3 + y^3 + 1 = 0 \), so \( d(x^3 + y^3 + 1) = 0 \).

By the properties of the differential we obtain
\[ 3x^2dx + 3y^2dy = 0, \text{ i.e. } \frac{dx}{y^2} = -\frac{dy}{x^2} \text{ on } U_1 \cap U_2. \]

So far we have defined a differential form on \( U_1 \cup U_2 \). Now on \( U_3 \), \( Z \) may take value zero so we choose affine coordinates \( u = X/Y, v = Z/Y \) and so the equation of the curve becomes \( u^3 + 1 + v^3 = 0 \). The \((x, y)\)- and the \((u, v)\)-coordinates are related as follows:
\[ x = \frac{X}{Z} = \frac{X}{Y} \cdot \frac{Y}{Z} = \frac{u}{v}, \quad y = \frac{Y}{Z} = \frac{1}{v}, \]
\[ u = \frac{X}{Z} = \frac{X}{Y} \cdot \frac{Y}{Z}, \quad v = \frac{Z}{Y}. \]
Thus \(dx/y^2\) on \(U_1 \cap U_3\) can be written
\[
\frac{dx}{y^2} = v^2d\left(\frac{u}{v}\right) = \frac{v^2vdu - dv}{v^2} = vdu - udv,
\]
which is regular in \(U_3\). Similarly, on \(U_2 \cap U_3\) we have
\[
-\frac{dy}{x^2} = -\frac{v^2}{u^2}d\left(\frac{1}{v}\right) = \frac{1}{u^2}dv.
\]
Again it is regular in \(U_3\). In fact, the two expressions coincide:
\[
d(u^3 + 1 + v^3) = 0 \Rightarrow du = -\frac{v^2}{u^2}dv \Rightarrow vdu - udv = \left(-\frac{v^3}{u^2} - u\right)dv = \frac{1}{u^2}dv,
\]
where we used that \(-u^3 - v^3 = 1\) on \(C\).

Summarizing, we have constructed a regular differential form \(\omega \in \Omega[C]\) such that \(\omega = dx/y^2\) on \(U_1\), \(\omega = -dy/x^2\) on \(U_2\), and \(\omega = dv/u^2\) on \(U_3\).

3.3.2. The Canonical Class. Similar to the divisor of a rational function we can define the divisor of a differential form \((\omega)\) for \(\omega \in \Omega(C)\).

**Definition 3.53.** Let \(\omega \in \Omega(C)\) be a rational differential form on \(C\). Let \(C = U_1 \cup \cdots \cup U_k\) be an open cover and \(\omega = f_i dt_i\) on \(U_i\). Define the divisor of \(\omega\)
\[
(\omega) = \sum_{p \in C} v_p(f_i) p,
\]
where for every \(p \in C\) we choose \(i\) such that \(U_i\) contains \(p\).

Note that there is a bit of freedom in this definition: there could be several \(U_i\) containing \(p\). Let’s see that the definition does not depend on which \(U_i\) we choose. If \(p \in U_i \cap U_j\) then \(f_i dt_i = f_j dt_j\) in \(U_i \cap U_j\). This implies that \(f_i/f_j dt_i = dt_j\) for \(f_i, f_j\) regular and non-zero in \(U_i \cap U_j\). But then \((f_i/f_j) = 0\) on \(U_i \cap U_j\), and so \((f_i) = (f_j)\) on \(U_i \cap U_j\). Therefore, \(v_p(f_i) = v_p(f_j)\).

Now what does the class of \((\omega)\) in \(\text{Cl}(C) = \text{Div}(C)/\mathbb{P}(C)\) look like? By definition, \((\omega)\) and \((g \omega)\), for \(g \in \mathbb{K}(C)\), belong to the same class in \(\text{Cl}(C)\). We will show that the divisors of any two forms belong to the same class in \(\text{Cl}(C)\). Let \(\omega, \omega'\) be two rational differential forms on \(C\) and \(\{(U_i, f_i)\} | 1 \leq i \leq k\), \(\{(U'_i, f'_i)\} | 1 \leq i \leq k'\) the corresponding open sets and rational functions. We can, in fact, assume that \(k = k'\) and \(U_i = U'_i\); if not we can consider a common refinement of the two open coverings, i.e. the collection \(\{U_i\} | 1 \leq i \leq k\} \cup \{U'_i\} | 1 \leq i \leq k'\}\). Second, we can assume that \(\omega = f_i dt_i\) and \(\omega' = f'_i dt_i\) on \(U_i\) since we can change the local parameter from \(f_i\) to \(f'_i\) so that will result in replacing the rational function \(f'_i\) with another rational function. Finally, we have a collection of rational functions \(f_i/f'_i\) on \(U_i\), for \(1 \leq i \leq k\) such that \(f_i/f'_i = f_j/f'_j\) on \(U_i \cap U_j\). But if two rational functions coincide on an open set \(U_i \cap U_j\) they must come from a unique rational function on \(U_i \cup U_j\). This shows that there exists a rational function \(g \in \mathbb{K}(C)\) such that \(g = f_i/f'_i\) on \(U_i\) for every \(1 \leq i \leq k\). In other words, any two rational forms \(\omega, \omega'\) on \(C\) are proportional:
\[
\omega = g \omega', \quad \text{for some } g \in \mathbb{K}(C),
\]
i.e. \((\omega)\) and \((\omega')\) lie in the same equivalence class in \(\text{Cl}(C)\). This shows that all rational forms define a unique class in \(\text{Cl}(C)\). This class is called the canonical class on \(C\) and is denoted by \(K_C\) (or simply \(K\) if it is clear what curve we are working with).
EXAMPLE 3.54.

(a) Let \( C = \mathbb{P}^1 = \mathbb{A}^1 \cup \{ \infty \} \) be the affine line with homogeneous coordinates \((x : z)\) and a local parameter \( t \) at the origin and \( u = 1/t \) at infinity. Consider the form \( \omega_u = dt \) in the open set \( z \neq 0 \) and \( \omega_v = -du/uz^2 \) in the open set \( x \neq 0 \). Note that \( \omega_0 \) has a pole at \( \infty = (1 : 0) \) of order two. Hence, \( (\omega_0) = -2 \cdot \infty \) and the canonical class \( K_{\mathbb{P}^1} = [-2 \cdot \infty] \in \text{Cl}(\mathbb{P}^1) \).

Any rational form on \( \mathbb{P}^1 \) looks like \( \omega = f(t)dt = f(t)\omega_0 \) in \( z \neq 0 \). We have \( (\omega) = (f) - 2 \cdot \infty \), i.e. \( (\omega) \sim (\omega_0) \), so again we see that all rational forms define the same class in \( \text{Cl}(\mathbb{P}^1) \). In particular, notice that \( \text{deg} K_{\mathbb{P}^1} = -2 \).

(b) Let \( C \) be the curve from Example 3.52, i.e. \( C = \{ X^3 + Y^3 + Z^3 = 0 \} \subset \mathbb{P}^2 \).

We have seen that there is a differential form \( \omega_0 \) on \( C \) which is regular everywhere on \( C \). Moreover, we saw that it does not take value zero in any of the open sets where it is defined. Therefore, \( (\omega_0) = 0 \in \text{Div}(C) \). This implies that \( K_C = [0] \in \text{Cl}(C) \) and \( \text{deg} K_C = 0 \). We can also describe all regular forms on \( C \):

\[ \Omega[C] = \{ f \omega_0 \mid f \text{ is regular on } C \} = \{ c \omega_0 \mid c \in \mathbb{K} \} = \text{span}_\mathbb{K} \{ \omega_0 \} \]

In particular, \( \text{dim}_\mathbb{K} \Omega[C] = 1 \).

In turns out that \( \text{dim}_\mathbb{K} \Omega[C] = g \), the genus of the curve \( C \).

**Theorem 3.55.** Let \( C \subset \mathbb{P}^2 \) be a smooth projective curve with homogeneous equation \( F(X, Y, Z) = 0 \). Let \( f(x, y) = 0 \) be the affine equation of \( C \). Then

\[ \Omega[C] = \left\{ \frac{hdx}{\partial f/\partial y} \mid h \in \mathbb{K}[x, y], \deg h \leq \deg C - 3 \right\} \]

Here \( \frac{hdx}{\partial f/\partial y} \) is the representation of a rational form in the open set \( \partial f/\partial y \neq 0 \).

**Proof.** Let \( x = X/Z, y = Y/Z \) be the affine coordinates in \( Z \neq 0 \) and let \( u = X/Y, v = Z/Y \) be the affine coordinates in \( Y \neq 0 \). These coordinates are related by

\[ x = \frac{u}{v}, \quad y = \frac{1}{v}. \]

In these open sets the curve is given by \( f(x, y) = 0 \) in \( Z \neq 0 \) and by \( g(u, v) = v^n f(u/v, 1/v) = 0 \), where \( n = \deg C \). We consider the following open cover of \( C \):

\[ U_1 = \{ (x : y : 1) \mid \partial f/\partial y \neq 0 \}, \quad U_2 = \{ (x : y : 1) \mid \partial f/\partial x \neq 0 \}, \]

\[ U_3 = \{ (u : 1 : v) \mid \partial g/\partial v \neq 0 \}, \quad U_4 = \{ (u : 1 : v) \mid \partial g/\partial u \neq 0 \}. \]

The form \( \frac{dx}{\partial f/\partial y} \) is regular on \( U_1 \) and the form \( -\frac{dy}{\partial f/\partial x} \) is regular on \( U_2 \). Moreover, they coincide on the intersection \( U_1 \cap U_2 \). Indeed, since \( f = 0 \) on \( C \) we have

\[ 0 = df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy \Rightarrow dx = \frac{dy}{\partial f/\partial x} \text{ in } U_1 \cap U_2. \]

Similarly, the form \( -\frac{dy}{\partial g/\partial u} \) is regular on \( U_3 \) and the form \( \frac{dx}{\partial g/\partial v} \) is regular on \( U_4 \) and they coincide on \( U_3 \cap U_4 \). In fact, these four forms define a regular form on \( C \). We need to check that they agree on \( U_1 \cap U_3, U_1 \cap U_4, U_2 \cap U_3, \) and \( U_2 \cap U_4 \). From \( g(u, v) = v^n f(u/v, 1/v) \) we have

\[ \frac{\partial g}{\partial u} = v^n \frac{\partial f}{\partial x} \frac{dx}{du} = v^{n-1} \frac{\partial f}{\partial x} \quad \text{and} \quad \frac{dy}{\partial v} = d \left( \frac{1}{v} \right) = -\frac{dv}{v^2}. \]
Therefore, on $U_2 \cap U_4$ we have
\[ -\frac{dy}{\partial f/\partial x} = \frac{v^{n-3} dv}{\partial g/\partial u}. \]
The rest is similar.

Now for any polynomial $h(x, y)$ the form $h\omega_0$ is clearly regular in $U_1 \cup U_2$. On $U_3 \cup U_4$ we have $h(x, y) = h(u/v, 1/v)$, so if $\deg h \leq n - 3$ then $v^{n-3}h(u/v, 1/v)$ is a polynomial in $u, v$. Therefore, the form $h\omega_0$ is also regular in $U_3 \cup U_4$. This shows that
\[ \{ h\omega_0 \mid h \in \mathbb{K}[x, y], \deg h \leq n - 3 \} \subseteq \Omega[C]. \]

On the other hand, any rational form looks like $h\omega_0$ for some rational function $h \in \mathbb{K}(C)$. If $h$ is not a polynomial then $h\omega_0$ is not regular in $U_1 \cup U_2$ and if $\deg h > n - 3$ then $h\omega_0$ is not regular at the infinite points of $C$ (where $v = 0$). Therefore,
\[ \{ h\omega_0 \mid h \in \mathbb{K}[x, y], \deg h \leq n - 3 \} = \Omega[C]. \]

This theorem has an important corollary which provides an alternative way to define the genus of a curve.

**Theorem 3.56.** Let $C$ be a smooth projective curve in $\mathbb{P}^2$. Then
\[ \dim_{\mathbb{K}} \Omega[C] = g, \]
where $g$ is the genus of the curve $C$.

**Proof.** Let $n = \deg C$. By the previous theorem the following differential forms are regular and span $\Omega[C]$
\[ \mathcal{B} = \left\{ \frac{x^i y^j dx}{\partial f/\partial y} \mid i + j \leq n - 3, i \geq 0, i \geq 0 \right\}. \]

In fact, it is a basis for $\Omega[C]$. Indeed, if there exist $c_{ij} \in \mathbb{K}$ such that $\sum c_{ij} \frac{x^i y^j dx}{\partial f/\partial y}$ defines the zero form on $C$ then the rational function $h(x, y) = \sum c_{ij} \frac{x^i y^j}{\partial f/\partial y}$ is zero on the open set $\partial f/\partial y \neq 0$ and, hence, zero everywhere on $C$. But then $f(x, y)h(x, y)$ and $\deg h \leq n - 3$ so $h(x, y)$ is the zero polynomial. This implies that $c_{ij} = 0$ for any $i, j$, i.e. the set $\mathcal{B}$ is linearly independent.

It remains to notice that $\mathcal{B}$ has \( \binom{n-3+2}{2} = \frac{(n-1)(n-2)}{2} \) elements and $g = \frac{(n-1)(n-2)}{2}$ by the Plücker formula (see Theorem 2.98).

### 3.4. The Riemann–Roch Formula

Let $C$ be a smooth projective curve over $\mathbb{K}$ and $D$ a divisor on $C$. In Section 3.2.2 we defined the Riemann–Roch space $\mathcal{L}(D)$. The Riemann–Roch formula relates all the notions we introduced before: the dimension $\ell(D)$ of $\mathcal{L}(D)$, the degree of $D$, the genus, and the dimension of the space $\mathcal{L}(K_C - D)$, where $K_C$ is the canonical class of $C$. Here is the statement.

**Theorem 3.57.** (The Riemann–Roch formula) Let $C$ be a smooth projective curve over $\mathbb{K}$ and $D \in \text{Div}(C)$. Then
\[ \ell(D) - \ell(K_C - D) = \deg D - g + 1, \]
where $K_C$ is the canonical class and $g$ is the genus of the curve.
We will prove this theorem in some special cases, the proof of the general case is beyond the scope of our course.

In fact, when \( C = \mathbb{P}^1 \) and \( D \) is effective we have already proved the Riemann–Roch formula in Proposition 3.46. Indeed, \( \mathbb{P}^1 \) has genus zero and \( \ell(K_{\mathbb{P}^1} - D) = \ell(-2\infty - D) \). Note that the degree of \(-2\infty - D\) is negative and in Exercise 3.4 you will show that \( \ell(D) = 0 \) if \( \deg D < 0 \). Hence, \( \ell(K_{\mathbb{P}^1} - D) = 0 \). The Riemann–Roch formula becomes

\[
\ell(D) = \deg D + 1,
\]
as in Proposition 3.46. For arbitrary \( D \) on \( \mathbb{P}^1 \) you will prove the Riemann–Roch theorem in Exercise 3.5. The idea is the same, any \( D \) is equivalent to \( m \infty \) for some \( m \in \mathbb{Z} \). Consider the three cases \( m > 0, m = 0, \) and \( m < 0 \) separately.

Now we will deduce several important facts from the Riemann–Roch formula.

**Corollary 3.58.** Let \( C \) be a smooth projective curve over \( \mathbb{K} \). Then

\[
\deg K_C = 2g - 2 = -\chi(C).
\]

*(Recall that \( \chi(C) \) is the Euler characteristic of \( C \)).*

**Proof.** Put \( D = K_C \) in the Riemann–Roch formula. Then

\[
\ell(K_C) - \ell(0) = \deg K_C - g + 1.
\]

As we saw in the proof of Theorem 3.43, \( \ell(0) = 1 \). Also

\[
\mathcal{L}(K_C) = \{ f \in \mathbb{K}(C)^* \mid (f) + K_C \geq 0 \} \cup \{ 0 \} = \{ f \in \mathbb{K}(C)^* \mid (f\omega_0) \geq 0 \} \cup \{ 0 \},
\]

where \( \omega_0 \) is some rational form on \( C \). Remember that any rational form on \( C \) looks like \( f\omega_0 \) for some \( f \in \mathbb{K}(C) \). Also, the condition \( (f\omega_0) \geq 0 \) means that \( f\omega_0 \) is a regular form on \( C \). Thus, \( \mathcal{L}(K_C) \) is isomorphic (as a vector space) to the space of all regular forms \( \Omega[C] \). Therefore, \( \ell(K_C) = \dim_{\mathbb{K}} \Omega[C] = g \), by Theorem 3.56. Putting everything together we obtain \( \deg K_C = 2g - 2 = -\chi(C) \) (see Theorem 2.92).

The following corollary allows us to compute the dimension of \( \mathcal{L}(D) \) when the degree of \( D \) is sufficiently large.

**Corollary 3.59.** Let \( C \) be a smooth projective curve over \( \mathbb{K} \) and \( D \in \text{Div}(C) \). If \( \deg D \geq 2g - 1 \) then

\[
\ell(D) = \deg D - g + 1.
\]

**Proof.** If \( \deg D \geq 2g - 1 \) then

\[
\deg(K_C - D) = 2g - 2 - \deg D \leq 2g - 2 - (2g - 1) = -1 < 0.
\]

Hence, by Exercise 3.4, \( \ell(K_C - D) = 0 \), and by the Riemann–Roch formula

\[
\ell(D) = \deg D - g + 1.
\]

**Corollary 3.60.** Any smooth curve of \( g = 0 \) is isomorphic to \( \mathbb{P}^1 \), i.e. there exists a rational function \( f : C \to \mathbb{P}^1 \) which is onto and one-to-one.

**Proof.** Choose any \( p \in C \) and apply the Riemann–Roch formula for \( D = p \). We get

\[
\ell(p) - \ell(K_C - p) = 1 - 0 + 1 = 2 \quad \Rightarrow \quad \ell(p) \geq 2.
\]

This implies that the Riemann–Roch space \( \mathcal{L}(p) \) contains a non-constant rational function \( f \). Since \( (f) + p \geq 0 \) the function \( f \) has only one pole at \( p \) and this is
a simple pole. (If $f$ has no poles then $f$ is regular and, hence, a constant.) We thus have $f(p) = \infty$ and $f^{-1}(\infty) = p$. To show $f$ is a bijection let $q \in \mathbb{P}^1 \setminus \{\infty\}$ and consider the function $f - q \in \mathbb{K}(C)$. Remember that the degree of a principal divisor $(f - q)$ is zero and

$$(f - q) = (f - q)_0 - (f - q)_\infty = (f - q)_0 - p,$$

so there is a unique point $p' \in C$ such that $(f - q)_0 = p'$, i.e. $f(p') = q$. Therefore, $f$ is a bijection.

\[ \square \]

### 3.4.1. The case of cubics.
Recall that any divisor on $\mathbb{P}^1$ is equivalent to $m \infty$ where $m = \deg D$. The situation with cubics is similar, although slightly more complicated.

Let $C$ be a smooth cubic over $\mathbb{K}$ and fix any point $p_0 \in C$. We have the following.

**Proposition 3.61.** Any divisor $D$ on $C$ is equivalent to $p + mp_0$ for some $p \in C$ and $m = \deg D - 1$.

**Proof.** First assume that $D$ is effective. We use induction on $m = \deg D - 1$. If $m = 0$ then $D = p$ and we are done. For $m > 0$ we can write $D = D' + q$ for some $q \in C$ and effective $D'$. By the inductive hypothesis $D' \sim p' + (m - 1)p_0$ for some $p' \in C$ and, hence,

$$D = D' + q \sim p' + q + (m - 1)p_0.$$ 

Once we show that $p' + q \sim p + p_0$ for some $p \in C$ we are done. Consider the line $L_1$ containing $p'$ and $q$. It intersects $C$ at one more point (by Bézout’s theorem), call it $q'$. Now let $L_2$ be the line containing $q'$ and $p_0$. It intersects $C$ at one more point which we call $p$. Now let $l_1 = 0$ be an equation of $L_1$ and consider the rational function $f = l_1/l_2 \in \mathbb{K}(C)$. We have

$$(f) = (f)_0 - (f)_\infty = (p' + q + q') - (q' + p_0 + p) = (p' + q) - (p + p_0),$$

which shows that $p' + q \sim p + p_0$. Note that if $p' = q$ then we need $L_1$ to be the tangent line to $C$ at $p'$. Similarly, in the other cases when some of the above points coincide.

Now, in general, $D = D_+ - D_-$ for some effective divisors $D_+$ and $D_-$. By the previous case $D_+ \sim p_+ + m_+ p_0$ and $D_- \sim p_- + m_- p_0$ for some $p_+, p_- \in C$ and $m_+ = \deg D_+ - 1$, $m_- = \deg D_- - 1$. Then $D \sim p_+ - p_- + (m_+ - m_-)p_0$. It remains to show that $p_+ - p_- \sim p - p_0$ for some $p \in C$ which is similar to the previous argument. \[ \square \]

Let us now see what the Riemann–Roch formula for a smooth cubic $C$ looks like. We know that $C$ has genus one and since $\dim_{\mathbb{K}} \Omega[C] = g = 1$ there exists a regular form $\omega_0$ on $C$ which spans (over $\mathbb{K}$) the space of all regular forms. Then $(\omega_0) \geq 0$ and $\deg(\omega_0) = 2g - 2 = 0$ imply that $(\omega_0) = 0$. Therefore, $K_C = 0$. The Riemann–Roch formula becomes

$$\ell(D) - \ell(-D) \equiv \deg D.$$ 

Note that if $\deg D > 0$ then $\ell(-D) = 0$ by Exercise 3.4 and the Riemann–Roch simplifies to

$$\ell(D) = \deg D.$$
3. THE RIEMANN–ROCH THEOREM

Similarly, if \( \deg D < 0 \) then \( \ell(D) = 0 \) and the Riemann–Roch says \( \ell(-D) = \deg(-D) \), which is the same statement as above.

If \( \deg D = 0 \) then \( D \sim p - p_0 \) for some \( p \in C \) by Proposition 3.61 (remember \( p_0 \) represents a fixed point in \( C \)). Hence

\[
\mathcal{L}(D) \cong \mathcal{L}(p - p_0) = \{ f \in \mathbb{K}(C)^* : (f) + p - p_0 \geq 0 \} \cup \{0\}.
\]

The condition \( (f) + p - p_0 \geq 0 \) means that \( f \) must have a zero at \( p_0 \). Also the only point \( f \) may have a pole at is \( p \) and, in fact, \( f \) must have a (simple) pole at \( p \), otherwise it would be a constant (the zero constant). Thus, \( (f) = p_0 - p \). But this means that \( f : C \to \mathbb{P}^1 \) is an isomorphism (see the proof of Corollary 3.60), which is impossible as \( C \) has genus one and \( \mathbb{P}^1 \) has genus zero. Therefore no such \( f \in \mathbb{K}(C)^* \) exists and \( \mathcal{L}(D) = \{0\} \). Similarly, \( \mathcal{L}(-D) \cong \mathcal{L}(p_0 - p) = \{0\} \) and the Riemann–Roch theorem is trivial.

We have shown that in the case of cubics the Riemann–Roch formula is equivalent to the following statement.

**Theorem 3.62.** Let \( C \) be a smooth cubic over \( \mathbb{K} \) and \( D \in \text{Div}(C) \) such that \( \deg D > 0 \). Then \( \ell(D) = \deg D \).

**Proof.** First we show that \( \ell(D) \leq \deg D \). Recall that \( \ell(D) \leq \deg D + 1 \) (see Theorem 3.43). Assume \( \ell(D) = \deg D + 1 \). Then using the same argument as in the proof of Theorem 3.43 (induction on \( \deg D \)) we see that \( \ell(p) = 2 \). We already saw in the proof of Corollary 3.60 that this implies that \( C \) is isomorphic to \( \mathbb{P}^1 \), which is not the case for a smooth cubic. Therefore, \( \ell(D) \leq \deg D \).

Now, let’s show that \( \ell(D) \geq \deg D \). The idea is to construct sufficiently many linearly independent elements of \( \mathcal{L}(D) \). Recall that \( D \sim p + m p_0 \) for some \( p \in C \) and \( m = \deg D - 1 \geq 0 \). Therefore, all we need to show is \( \ell(p + m p_0) \geq m + 1 \) for any \( m \geq 0 \).

We use induction on \( m \). If \( m = 0 \) then \( \ell(p) \geq 1 \) since \( \mathcal{L}(p) \) contains constants. Suppose \( m > 0 \). First, we will look at three cases: \( m = 1, 2, 3 \).

**Case** \( m = 1 \). Let \( L_2 \) be the line containing \( p, p_0 \) and let \( q \) be the third intersection point of \( L_2 \) with \( C \). Let \( L_1 \) be any line containing \( q \) and let \( p_1, p_2 \) be the other intersection points of \( L_1 \) with \( C \). Consider \( f = l_1/l_2 \in \mathbb{K}(C) \), where \( l_1, l_2 \) are linear polynomials defining the lines \( L_1, L_2 \). Then \( (f) + p + p_0 = p_1 + p_2 \geq 0 \), so \( \mathcal{L}(p + p_0) \) contains a non-constant function \( f \), hence, \( \ell(p + p_0) \geq 2 \).

**Case** \( m = 2 \). First, note that \( \mathcal{L}(p + p_0) \subseteq \mathcal{L}(p + 2 p_0) \), so if we show that there exists \( f_2 \in \mathcal{L}(p + 2 p_0) \) such that \( f_2 \notin \mathcal{L}(p + p_0) \) this will guarantee that \( \ell(p + 2 p_0) > \ell(p + p_0) \geq 2 \) and so \( \ell(p + 2 p_0) \geq 3 \), as required. In fact, we will show that there exists \( f_2 \in \mathcal{L}(p + 2 p_0) \) with \( (f_2)_\infty = 2 p_0 \). Let \( L_2 \) be the tangent line to \( C \) at \( p_0 \) and let \( q \) be the third intersection point of \( L_2 \) with \( C \). As before, let \( L_1 \) be any line containing \( q \) and let \( p_1, p_2 \) be the other intersection points of \( L_1 \) with \( C \). Then \( f_2 = l_1/l_2 \in \mathbb{K}(C) \), where \( l_1, l_2 \) are linear polynomials defining the lines \( L_1, L_2 \), satisfies

\[
(f_2) + p + 2 p_0 = p + p_1 + p_2 \geq 0, \quad \text{and} \quad (f_2)_\infty = 2 p_0.
\]

**Case** \( m = 3 \). The idea is the same as in Case \( m = 2 \): We want to construct a function \( f_3 \in \mathcal{L}(p + 3 p_0) \) with \( (f_3)_\infty = 3 p_0 \). We put \( f_3 = f_2 l_3/l_4 \), where \( l_4 \) is a linear polynomial defining the line containing \( p_0, p_2 \) and one more point \( q' \in C \) and \( l_3 \) is a linear polynomial defining a line containing \( q' \) and two more points.
\(q_1, q_2 \in C\). Then
\[
(f_3) + p + 3p_0 = p + p_1 + q_1 + q_2 \geq 0, \quad \text{and} \quad (f_3)_\infty = 3p_0.
\]

Finally, for any \(m > 0\) there exists \(f_m \in \mathcal{L}(p + m p_0)\) with \((f_m)_\infty = m p_0\). Indeed, if \(m = 2r\) take \(f_m = f_2^r\) and if \(m = 2r + 3\) take \(f_m = f_2^r f_3\). Therefore, 
\[
\ell(p + m p_0) > \ell(p + (m - 1) p_0).
\]
By the inductive hypothesis \(\ell(p + (m - 1) p_0) \geq m\), so \(\ell(p + m p_0) \geq m + 1\), as required. \(\square\)

### 3.4.2. Special divisors and Weierstrass points.

We are back to the general case when \(C\) is a smooth projective curve over \(\mathbb{K}\).

**Definition 3.63.** A divisor \(D \in \text{Div}(C)\) is called **special** if \(\ell(K_C - D) > 0\).

In Exercise 3.7 you will show that if \(\deg D \leq g - 2\) then \(D\) is special and if \(\deg D \geq 2g - 1\) then \(D\) is not special.

We will be dealing with divisors supported on just one point, \(D = a p\), for some \(p \in C\) and positive integer \(a\). Our goal is to understand the function \(\ell(a p)\) as a function of \(a \in \mathbb{Z}, \ a \geq 1\).

First, note that this is an increasing function as \(\mathcal{L}(D) \subseteq \mathcal{L}(D + p)\) for any divisor \(D\). Also, it may increase by at most one at every step. Indeed, we have
\[
\mathcal{L}(K_C - a p) \subseteq \mathcal{L}(K - a p + p), \quad \text{hence} \quad \ell(K_C - a p) \leq \ell(K_C - (a - 1)p).
\]
Therefore, applying the Riemann–Roch formula twice we obtain
\[
\ell(a p) = \ell(K_C - a p) + a - g + 1 \leq \ell(K_C - (a - 1)p) + a - g + 1
\]
\[
= (\ell((a - 1)p) - (a - 1) + g - 1) + a - g + 1 = \ell((a - 1)p) + 1.
\]
This shows that
\[
\text{either} \quad \ell(a p) = \ell((a - 1)p) \quad \text{or} \quad \ell(a p) = \ell((a - 1)p) + 1.
\]

**Definition 3.64.** We say that \(a \geq 1\) is a gap at \(p \in C\) if \(\ell(a p) = \ell((a - 1)p)\), and a non-gap otherwise.

In other words, \(a\) is a non-gap if and only if there exists \(f \in \mathbb{K}(C)\) with pole of order \(a\) at \(p\) and no other poles on \(C\), as this would imply that \(\mathcal{L}((a - 1)p) \subseteq \mathcal{L}(a p)\).

Note that if \(C\) is a smooth cubic then all \(a \geq 1\) are non-gaps, since \(\ell(a p) = a\) by Theorem 3.62. Thus, we are going to assume that \(C\) has genus \(g \geq 2\).

The gaps at \(p \in C\) satisfy the following properties.

**Proposition 3.65.** *Let \(C\) be a smooth curve of genus \(g \geq 2\) and \(p \in C\).*

1. If \(a, b\) are non-gaps at \(p\) then so is \(a + b\).
2. 1 is a gap at \(p\).
3. If \(a \geq 2g\) then \(a\) is a non-gap at \(p\).
4. The number of gaps at \(p\) equals \(g\).

**Proof.** (1) If there exists \(f, g \in \mathbb{K}(C)\) with \((f)_\infty = a p\) and \((g)_\infty = b p\) then \(fg\) satisfies \((fg)_\infty = (a + b) p\).

(2) We have \(\ell(p) = \ell(0) = 1\). Indeed, if \(\ell(p) \geq 2\) then there exists a non-constant \(f \in \mathbb{K}(C)\) with \((f)_\infty = p\), which implies that \(f : C \rightarrow \mathbb{P}^1\) is an isomorphism (see the proof of Corollary 3.60). This is impossible since \(g \geq 2\).

(3) If \(a \geq 2g\) then the divisors \((a - 1)p\) and \(a p\) are not special, by Exercise 3.7. Then the Riemann–Roch formula produces \(\ell((a - 1)p) = a - 1 - g + 1\) and \(\ell(a p) = a - g + 1\), i.e. \(a\) is a non-gap at \(p\). In particular, \(\ell(2g p) = g + 1\).
(4) Think about the graph of \( \ell(a p) \) on the segment \( 0 \leq a \leq 2g \). It takes value 1 at \( a = 0 \) and value \( g + 1 \) at \( a = 2g \). Hence the graph of \( \ell(a p) \) goes up by \( g \) steps and to the right by \( 2g \) steps. There must be exactly \( 2g - g = g \) places where it does not go up.

We will call the sequence of gaps \( a_1 \prec a_2 \prec \cdots \prec a_g \) at \( p \in C \), the gap sequence.

**Definition 3.66.** A point \( p \in C \) is called a Weierstrass point if the gap sequence \( a_1 \prec a_2 \prec \cdots \prec a_g \) at \( p \) does not coincide with \( 1 \prec 2 \prec \cdots \prec g \).

When \( p \) is not Weierstrass, the graph of \( \ell(a p) \) looks particularly simple.

![Figure 3.1. The graph of \( \ell(a p) \) when \( p \) is not a Weierstrass point.](image)

**Proposition 3.67.** The following are equivalent.

1. \( p \) is a Weierstrass point.
2. \( D = gp \) is a special divisor.
3. \( D = ap \) is a special divisor for some \( a \geq g \).

**Proof.** (1) \( \Rightarrow \) (2) Assume \( gp \) is not a special divisor. Then \( \ell(K_C - gp) = 0 \), so by the Riemann–Roch formula \( \ell(gp) = g - g + 1 = 1 \). This implies that \( \ell(a p) = 1 \) for all \( 0 \leq a \leq g \) (as \( \ell(a p) \) is increasing and \( \ell(0) = 1 \)). By definition, \( p \) is not Weierstrass.

(2) \( \Rightarrow \) (3) This is trivial, just take \( a = g \).

(3) \( \Rightarrow \) (1) Suppose \( ap \) is special for some \( a \geq g \). Then \( \ell(K_C - ap) > 0 \) and so \( \ell(a p) > a - g + 1 \), by the Riemann–Roch formula. But if \( p \) is not Weierstrass then \( \ell(a p) = a - g + 1 \) on \( a \geq g \) (see Figure 3.1). Thus \( p \) is Weierstrass.

With every point \( p \in C \) we can associate its weight, which by definition equals

\[
\text{w}(p) = \sum_{i=1}^{g}(a_i - i),
\]

where \( a_1 \prec a_2 \prec \cdots \prec a_g \) is the gap sequence at \( p \). Note that \( p \) is Weierstrass if and only if \( \text{w}(p) > 0 \). In Exercise 3.8 you will use the properties of the gaps to show that if 2 is a non-gap at \( p \) then \( \text{w}(p) = \binom{g}{2} \).
3.5. Elliptic Curves

Let \( C \) be a smooth cubic in \( \mathbb{P}^2 \) and fix a point \( p_0 \in C \). Recall that every divisor \( D \) on \( C \) satisfies \( D \sim p + m p_0 \) for some \( p \in C \) and \( m = \deg D - 1 \) (see Proposition 3.61). This allows us to define a bijection between points \( p \) of \( C \) and the subgroup \( \text{Cl}^0(C) \subset \text{Cl}(C) \) of divisor classes of degree zero.

**Theorem 3.68.** Let \( \text{Cl}^0(C) \subset \text{Cl}(C) \) be the subgroup of divisor classes of degree zero. The map

\[
\phi : C \to \text{Cl}^0(C), \quad p \mapsto [p - p_0]
\]

is a bijection.

**Proof.** The fact that \( \phi \) is onto follows from Proposition 3.61: any divisor \( D \) of degree zero is equivalent to \( p - p_0 \) for some \( p \in C \) (since \( m = \deg D - 1 = -1 \)).

To show that \( \phi \) is one-to-one suppose \( p - p_0 \sim q - p_0 \) for some \( p \hookrightarrow q \in C \). We claim that \( p = q \). Indeed, \( p - q = (f) \). But if \( p \neq q \) then \( f \) defines an isomorphism \( f : C \to \mathbb{P}^1 \) (remember the proof of Corollary 3.60). This is a contradiction as \( C \) has genus 1. \( \square \)

### 3.5.1. The Group Law

Now since \( \text{Cl}^0(C) \) is an Abelian group we should be able to translate the group structure from \( \text{Cl}^0(C) \) to \( C \) itself. The idea is that if \( \phi(p) = [p - p_0] \) and \( \phi(q) = [q - p_0] \) then there exists a unique point in \( C \), which we denote by \( p \oplus q \), such that \( \phi(p \oplus q) = [p + q - 2 p_0] \). Indeed, by Proposition 3.61, \( p + q - 2 p_0 \sim p' - p_0 \) for some \( p' \in C \). This \( p' \) is the point \( p \oplus q \) we are looking for.

Next we will see how to describe \( p' = p \oplus q \) geometrically. The condition \( p + q - 2 p_0 \sim p' - p_0 \) is equivalent to \( p + q \sim p' + p_0 \). Let \( L_1 \) be the line containing \( p, q \) and let \( q' \) be the third intersection point of \( L_1 \) and \( C \). Let \( L_2 \) be the line containing \( q', p_0 \) and let \( p' \) be the third intersection point of \( L_2 \) and \( C \). Then the function \( f = l_1/l_2 \), where \( l_i \) is the linear polynomial defining \( L_i \), satisfies

\[
(f) = p + q + q' - q' - p_0 - p', \quad \text{i.e.} \quad p + q = (f) + p' + p_0,
\]

which means that \( p + q \sim p' + p_0 \).

![Figure 3.2](image-url)  
**Figure 3.2.** The construction of \( p \oplus q \). The fixed point \( p_0 \) is the origin.
3.5.2. Weierstrass normal form. It turns out that after a suitable (projective) change of coordinates a smooth cubic can be written as \( y^2 = x^3 + ax + b \) for some \( a, b \in \mathbb{K} \). This is so called the Weierstrass normal form.

Recall that a projective change of coordinates in \( \mathbb{P}^2 \) is given by a \( 3 \times 3 \) matrix \( A \in GL(3, \mathbb{K}) \):

\[
\begin{pmatrix}
X \\
Y \\
Z
\end{pmatrix} =
\begin{pmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{pmatrix}
\begin{pmatrix}
U \\
V \\
W
\end{pmatrix},
\]

where \((X : Y : Z)\) and \((U : V : W)\) are the homogeneous coordinates on \( \mathbb{P}^2 \). In the affine coordinates \( x = X/Z, y = Y/Z, \) and \( u = U/W, v = V/W \) this corresponds to the following rational transformation:

\[
x = \frac{a_{11}u + a_{12}v + a_{13}}{a_{31}u + a_{32}v + a_{33}}, \quad y = \frac{a_{21}u + a_{22}v + a_{23}}{a_{31}u + a_{32}v + a_{33}}.
\]

**Theorem 3.69.** (Weierstrass normal form) Let \( F(X, Y, Z) = 0 \) be an equation of a smooth cubic. If \( \text{char} \mathbb{K} \neq 2, 3 \) then there is a projective change of coordinates in which \( F(X, Y, Z) = 0 \) becomes \( V^2W = U^3 + aUW^2 + bW^3 \) for some \( a, b \in \mathbb{K} \).

In affine coordinates \( u^2 = a^3 + au + b \).

The proof of this is, although elementary, rather technical, so we do not include it here. You can find it in [ ]. You will compute the Weierstrass normal form of the Fermat curve \( x^3 + y^3 = 1 \) in Exercise 3.9, where all the necessary steps are outlined.

Note that all the smooth cubics we depicted previously were in the Weierstrass normal form.

From now on we will assume that a smooth cubic \( C \) is given by the affine equation \( y^2 = x^3 + ax + b \). The infinite point \( p_0 = (0 : 1 : 0) \) lies on \( C \). Moreover, the tangent line to \( C \) at \( p_0 \) has local intersection number three with \( C \) at \( p_0 \), i.e. \( p_0 \) is an inflection point of \( C \). Indeed, the cubic is defined by the homogeneous polynomial \( F(X, Y, Z) = X^3 + aXZ^2 + bZ^3 - Y^2Z \) and the tangent line \( E \) has parametric equation \( X = t, Y = 1, Z = 0 \). The polynomial \( F(t, 1, 0) = t^3 \) has \( t = 0 \) as a root of multiplicity three, which equals the intersection number \((C \cdot E)_{p_0}\) by Exercise 2.13.

Now let us get back to the group law on \( C \). We choose \( p_0 = (0 : 1 : 0) \in C \). Then given \( p, q \in C \), here is how we construct the point \( p \oplus q \) geometrically. First, find \( q' \in C \) which lies on the line joining \( p \) and \( q \). Since lines through \( p_0 \) are the vertical lines, we let \( p \oplus q \) be the point on the vertical line containing \( q' \)(see Figure 3.3 for the group law on the real part of \( C \)).

**Example 3.70.**

(a) What is the zero element of the group, i.e. the point \( 0 \in C \) such that \( p \oplus 0 = p \) for any \( p \in C \)? By construction \( p = p' \) means that \( L_1 \) coincides with \( L_2 \), hence \( L_1 \) is vertical and \( q = p_0 \). Therefore, \( 0 = p_0 = (0 : 1 : 0) \).

(b) What is \( 2p = p \oplus p' \)? In this case \( p = q \), so \( L_1 \) is the tangent line to \( C \) at \( p \). See Figure 3.4 illustrating this.

(c) What is the inverse \( -p \) of \( p \)? By definition \( -p \) is the point such that \( p \oplus (-p) = 0 \). In other words, \( p' = p_0 \) which means that \( L_2 \) is tangent to \( C \) at \( p_0 \). We have already seen that \( p_0 \) is an inflection point of \( C \), so \( q' \) must also equal \( p_0 \). This shows that \( L_1 \) is vertical and \( -p \) is the reflection of \( p \) about the \( x \)-axis (see Figure 3.5).
3.5. ELLIPTIC CURVES

Figure 3.3. The construction of $p \oplus q$. The fixed point $p_0$ is $(0 : 1 : 0)$.

Figure 3.4. The construction of $2p = p \oplus p$.

(d) Let $L_1$ be any line and $C \cap L_1 = \{p, q, q'\}$. Then $p \oplus q \oplus q' = 0$. Indeed, by construction $p \oplus q$ is the reflection of $q'$ about the $x$-axis. By the previous example this means that $p \oplus q = -q'$, i.e. $p \oplus q \oplus q' = 0$.

So far we have shown that all of the group axioms hold on $C$, except for associativity. We will prove it geometrically below. Note that the group law on $C$ is clearly commutative.

**Proposition 3.71.** The operation $p \oplus q$ on $C$ is associative.

**Proof.** We need to show that for any $p, q, r \in C$ we have $(p \oplus q) \oplus r = p \oplus (q \oplus r)$. Let $L_1, L_2$ be the lines appearing in the construction of $p \oplus q$ and, similarly, $M_1, M_2$ be the lines appearing in the construction of $q \oplus r$. Furthermore, let $L_3$ be the line
containing $p$ and $q \oplus r$ and $M_3$ be the line containing $r$ and $p \oplus q$. Let $a$ be the intersection point of $L_3$ and $M_3$. If we show that $a$ lies on $C$ then we are done, since on one hand $a = -(p \oplus (q \oplus r))$ (as $L_3 \cap C = \{p, q \oplus r, a\}$) and on the other hand $a = -((p \oplus q) \oplus r)$ (as $M_3 \cap C = \{p \oplus q, r, a\}$).

To show $a \in C$ we apply Chasles’ theorem. The two cubics $L_1 \cup M_2 \cup M_3$ and $M_1 \cup L_2 \cup L_3$ intersect in nine points $\{0, p, q, r, \pm(p \oplus q), \pm(q \oplus r), a\}$. The cubic $C$ contains the first eight of them, hence, by Chasles’ theorem must contain $a$ as well. \qed
3.5.3. Elliptic curves over \( \mathbb{Q} \). We close this chapter with discussion of elliptic curves given by polynomials with rational coefficients. We will call them rational elliptic curves (not to be confused with rational curves as in Definition 2.74, in that sense they are not rational as they have genus one). They play an important role in various applications such as in number theory and cryptography. Despite the simpleness of the definition, rational elliptic curves hide many open questions. We will discuss only a few.

Definition 3.72. A smooth cubic \( C \) which is the zero set of a homogeneous polynomial \( F \in \mathbb{Q}[X,Y,Z] \) is called a rational elliptic curve.

If \( C \) has a rational point (i.e. a point with rational coordinates) then after a rational projective transformation (i.e. whose matrix lies in \( \text{GL}(3, \mathbb{Q}) \)) we can bring \( C \) to the Weierstrass normal form \( Y^2Z = X^3 + aXZ^2 + bZ^3 \) with \( a, b \in \mathbb{Q} \).

Proposition 3.73. Let \( p, q \in C \) be rational points on \( C \). Then \( p \oplus q \) is also rational.

Proof. If \( p \) or \( q \) is \( 0 = (0 : 1 : 0) \) or if \( p \oplus q = 0 \), the statement is trivial. Hence we may assume that \( p, q \), and \( p \oplus q \) lie in the affine part of \( C \). Let \( p = (x_p : y_p : 1) \) and \( q = (x_q : y_q : 1) \) and consider \( q' = (x_{q'} : y_{q'} : 1) \) which is the third intersection point of \( C \) with the line containing \( p, q \) (see Figure 3.3). Note that \( p \oplus q = (x_{q'} : -y_{q'} : 1) \) so we need to show that \( x_{q'}, y_{q'} \in \mathbb{Q} \).

To find \( x_{q'} \) and \( y_{q'} \) we need to solve the system of two (affine) equations:

\[
y^2 = x^3 + ax + b, \quad (x - x_p)(y_q - y_p) = (y - y_p)(x_q - x_p).
\]

Since \( x_p \neq x_q \) (otherwise \( p \oplus q = 0 \), see Figure 3.5) we can express \( y \) from the second equation

\[
y = y_p + \lambda(x - x_p), \quad \text{where } \lambda = \frac{y_q - y_p}{x_q - x_p} \text{ is the slope.}
\]

Clearly \( \lambda \in \mathbb{Q} \) since \( p, q \) have rational coordinates. Now, plugging this expression into the first equation, we obtain a cubic equation in \( x \)

\[
(y_p + \lambda(x - x_p))^2 = x^3 + ax + b,
\]

whose roots are \( x_p, x_q, \) and \( x_{q'} \). Since the sum of the roots is the negative of the coefficient of \( x^2 \) we obtain

\[
x_p + x_q + x_{q'} = \lambda^2,
\]

hence,

\[
x_{q'} = \lambda^2 - x_p - x_q, \quad y_{q'} = y_p + \lambda(\lambda^2 - 2x_p - x_q).
\]

This shows that \( x_{q'}, y_{q'} \in \mathbb{Q} \). \( \square \)

Corollary 3.74. The set \( C(\mathbb{Q}) \) of all rational points on \( C \) is a group with respect to the operation \( \oplus \).

Rational elliptic curves \( C \) attracted a lot of attention in the 20th century mathematics, in particular because of their connection with Fermat’s Last Theorem. Below we state two famous theorems about the structure of the group \( C(\mathbb{Q}) \). Ideas appearing in the proofs of these theorems were among the key ingredients of Andrew Wiles’s proof of Fermat’s Last Theorem (1993). We will not need these theorems in the remaining of our exposition.
Definition 3.75. A point \( p \in C \) is called a torsion point if there exists \( m \geq 1 \) such that

\[
mp := p \oplus p \oplus \cdots \oplus p = 0.
\]

The smallest such \( m \) is called the order of \( p \) in \( C \). If \( p \) is not a torsion point, we say it has infinite order.

The following proposition you should check yourself.

Proposition 3.76. Let \( C \) be a rational elliptic curve.

1. If \( p \hookrightarrow q \in C \) are torsion points then so is \( p \oplus q \).
2. The size of the cyclic subgroup \( \langle p \rangle := \{ mp \mid m \in \mathbb{Z} \} \) equals the order of \( p \) in \( C \).

Note, for \( m < 0 \) the expression \( mp \) denotes \( (-m)(-p) \).

Corollary 3.77. The set

\[
C_{\text{tor}}(\mathbb{Q}) = \{ \text{torsion points in } C \}
\]

forms a subgroup of \( C(\mathbb{Q}) \) with respect to the operation \( \oplus \).

Since \( C_{\text{tor}}(\mathbb{Q}) \) is a finite abelian group it is isomorphic to the direct product of finite cyclic groups (see, for example, Sec. ?? in \[\text{?}\]). In the theorem below Barry Mazur described all possible groups \( C_{\text{tor}}(\mathbb{Q}) \) up to isomorphism.

Theorem 3.78. (Mazur, 1977) Let \( C \) be a rational elliptic curve. The group \( C_{\text{tor}}(\mathbb{Q}) \) is isomorphic to one of the following:

\[
\mathbb{Z}_n, \quad \text{for } n = 2, 3, \ldots, 10, 12, \quad \text{and} \quad \mathbb{Z}_2 \times \mathbb{Z}_n, \quad \text{for } n = 2, 4, 6, 8.
\]

The following result by Louis Joel Mordell asserts the existence of a finite “basis” for the group \( C(\mathbb{Q}) \). Let us first give a definition.

Definition 3.79. Let \( p_1, \ldots, p_r \) be points of \( C \). We say that they are linearly independent if the points \( m_1p_1 \oplus \cdots \oplus m_rp_r \) for \( m_i \in \mathbb{Z} \) are all distinct.

Theorem 3.80. (Mordell, 1922) Let \( C \) be a rational elliptic curve. There exist linearly independent \( p_1, \ldots, p_r \in C(\mathbb{Q}) \) such that every point \( p \in C(\mathbb{Q}) \) has a form

\[
p = a_1p_1 \oplus \cdots \oplus a_r p_r \oplus q
\]

for unique \( a_i \in \mathbb{Z} \) and \( q \in C_{\text{tor}}(\mathbb{Q}) \).

Definition 3.81. The smallest \( r \) satisfying Mordell’s theorem is called the rank of \( C \).

The famous Rank Problem asks: Can the rank can be arbitrarily large? This is an open question, so far elliptic curves of rank up to 24 have been found.
Exercises

Exercise 3.1. Let \( C \) be an affine irreducible curve over \( \mathbb{F} \). Suppose \( g \in \mathbb{F}(C) \) is regular at \( p \in C \). Define the value of \( g \) at \( p \) as \( g(p) = a(p)/b(p) \), where \( g = a/b \) is a representation of \( g \) such that \( a, b \in \mathbb{F}[x, y] \) and \( b(p) \neq 0 \). Show that \( g(p) \) is independent of the representation of \( g \) in \( \mathbb{F}(C) \).

Exercise 3.2. Let \( C \) be a cubic defined by \( y^2 = x^3 \). Show that neither \( y/x \) nor \( x/y \) defines a regular function at \( (0,0) \).

Exercise 3.3. Let \( C \) be the smooth cubic \( y^2z = x^3 - xz^2 \). Show that \( dx/y \) extends to a regular differential form on \( C \).

Exercise 3.4. Let \( D \) be a divisor on \( C \) with \( \deg(D) < 0 \). Describe \( L(D) \) and compute \( \ell(D) \). (Hint: What is the degree of the divisor \((f) + D\) if \((f) + D \geq 0\)?)

Exercise 3.5. Prove the Riemann–Roch theorem on \( \mathbb{P}^1 \): For any divisor \( D \) on \( \mathbb{P}^1 \) we have

\[
\ell(D) - \ell(-2\infty - D) = \deg(D) + 1.
\]

(Note: This statement was proved in Proposition 3.46 for effective divisors.)

Exercise 3.6. Recall that a divisor \( D \) is called special if \( \ell(K - D) > 0 \). Prove that \( D \) is special if and only if \( D \sim K - D' \) for some effective divisor \( D' \).

Exercise 3.7. Let \( D \) be any divisor on \( C \). Prove the following:

(a) If \( \deg(D) \geq 2g - 1 \) then \( D \) is not special.

(b) If \( \deg(D) \leq g - 2 \) then \( D \) is special. (Hint: Use the Riemann–Roch theorem.)

Here \( g \) denotes the genus of \( C \).

Exercise 3.8. Let \( C \) be a smooth projective curve of genus \( g \) and \( p \in C \). Assume 2 is a non-gap at \( p \). Show that the Weierstrass weight of \( p \) satisfies \( w(p) = \left(\frac{g}{2}\right) \). (Hint: Can you identify the gap sequence in this case?)

Exercise 3.9. Compute the Weierstrass normal form for the Fermat curve \( x^3 + y^3 = 1 \) using the following steps:

(a) Homogenize the above equation.

(b) Apply the projective transformation

\[
\begin{pmatrix}
X \\
Y \\
Z
\end{pmatrix} = \begin{pmatrix}
0 & -6 & 1 \\
0 & 6 & 1 \\
6 & 0 & 0
\end{pmatrix} \begin{pmatrix}
U \\
V \\
W
\end{pmatrix}
\]

(c) Dehomogenize the equation to get \( v^2 = u^3 + au + b \) for some rational \( a, b \).

Exercise 3.10. Let \( p = (x_p, y_p) \) be a point on an elliptic curve \( y^2 = x^3 + ax + b \). Compute the coordinates of \( p \oplus p \). They should depend rationally on \( x_p \) and \( y_p \).
EXERCISE 3.11. Consider an elliptic curve $C$ given by $y^2 = x^3 + 1$ and let $0 = (0 : 1 : 0)$.

(a) Let $p \in C$ and $L$ a line such that $L \cap C = \{p, p, q\}$ (i.e. $L$ is tangent to $C$ at $p$). Explain how to construct $p + q$.
(b) Let $p \in C$ has zero $y$-coordinate. What is $p \oplus p$?
(c) Let $p = (0 : 1 : 1) \in C$. Show that $p$ is an inflection point of $C$. Explain how to construct $p \oplus p$.

EXERCISE 3.12. Consider an elliptic curve $y^2 = x^3 + ax + b$. Explain how to compute the coordinates of the points of order two on the curve $y^2 = x^3 + ax + b$. (Recall that a point $p$ has order two if $p \oplus p = 0$.)

EXERCISE 3.13. The points $p$ of order 3 on $y^2 = x^3 + ax + b$ are the inflection points, where $y'' = 0$. Find the x-coordinates of the points of order 3 on $y^2 = x^3 - x$. How many real and how many complex points of order 3 are there?
CHAPTER 4

Curves over Finite Fields

Up to the end of the previous chapter we have been working with curves over algebraically closed fields. For the purpose of coding theory however we need to consider curves over finite fields, which are not algebraically closed. (Can you show that algebraically closed fields are infinite?) The central result of this chapter is the Hasse–Weil theorem which provides a bound for the number of solutions to polynomial equations over finite fields. Although we don’t have enough tools to prove the theorem (this alone could be the topic of a course) we will connect this theorem to a version of the Riemann Hypothesis for curves over finite fields.

4.1. Curves over non-algebraically closed fields

We begin by defining curves over arbitrary fields and seeing how we should adjust our definitions and theorem in the case of non-algebraically closed field.

Let $K$ be any field and $\overline{K}$ be its algebraic closure.

**Definition 4.1.** We say that $C \subset \mathbb{P}^2_K$ is a curve over $K$ if its defining equations has coefficients in $K$, i.e.

$$C = \{(x : y : z) \in \mathbb{P}^2_K \mid F(x, y, z) = 0\}$$

for some homogeneous polynomial $F \in K[x, y, z]$.

We will be interested in points of $C$ whose coordinates also lie in $K$ (or in a finite extension of $K$). Such points will be called $K$-rational points (or simply $K$-points of $C$) and denoted by $C(K)$. In other words,

$$C(K) = \{(x : y : z) \in \mathbb{P}^2_K \mid F(x, y, z) = 0\}.$$

**Example 4.2.**

(a) Rational elliptic curves from Section 3.5.3 are curves over $\mathbb{Q}$ and $C(\mathbb{Q})$ is the set of rational points of $C$.

(b) The unit circle $C = \{(x : y : z) \in \mathbb{P}_C \mid x^2 + y^2 = z^2\}$ can be considered as a curve over $\mathbb{Q}$ or over any extension of $\mathbb{Q}$. We thus have $(3 : 4 : 5) \in C(\mathbb{Q})$, $(\sqrt{3} : 1 : 2) \in C(\mathbb{Q}(\sqrt{3}))$, and $(i : \sqrt{2} : 1) \in C(\mathbb{C})$ (or we can say $(i : \sqrt{2} : 1) \in C(\mathbb{Q}(i, \sqrt{2}))$). Note that “conjugates” points $(-\sqrt{3} : 1 : 2)$ and $(-i : \sqrt{2} : 1)$ also lie on the circle $C$. We state this in general in the following proposition.

**Proposition 4.3.** Let $K \subset K'$ be a finite extension and $C$ a curve over $K$. Then for any $\phi \in \text{Gal}(K'/K)$ and any $p \in C$ we have $p \in C(K')$ if and only if $\phi(p) \in C(K')$.

**Proof.** Suppose $C$ is given by $F(x, y, z) = 0$ for a homogeneous polynomial $F \in K[x, y, z]$ and let $p = (x_p : y_p : z_p)$ be in $C(K')$. Then $F(x_p, y_p, z_p) = 0$ so

$$\phi(F(x_p, y_p, z_p)) = \phi(0) = 0.$$
On the other hand,
\[
\phi(F(x_p, y_p, z_p)) = F(\phi(x_p), \phi(y_p), \phi(z_p))
\]
since the coefficients of \(F\) are fixed by \(\phi\). Therefore \(F(\phi(x_p), \phi(y_p), \phi(z_p)) = 0\), i.e.,
\[
\phi(p) = (\phi(x_p) : \phi(y_p) : \phi(z_p)) \text{ lies in } C(K').
\]
The converse is the same by replacing \(\phi\) with its inverse \(\phi^{-1}\).

It is convenient to consider the set of points \(\phi(p)\) for all \(\phi \in \text{Gal}(K'/K)\) as one “point” of \(C\). More explicitly we have the following definition.

**Definition 4.4.** Let \(C\) be a curve over \(K\) and \(K \subset K'\) a finite field extension. A set
\[
\{\phi(p) \mid \phi \in \text{Gal}(K'/K)\}
\]
for some \(p \in C(K')\) is called a **point of \(C\) of degree** \(|\text{Gal}(K'/K)|\).

In some literature a point in \(C\) is called a **place** in \(C\).

Coming back to our previous example, the set \(((\sqrt{3} : 1 : 2), (-\sqrt{3} : 1 : 2))\) is a point of \(x^2 + y^2 = z^2\) of degree two. Here \(K = \mathbb{Q}\) and \(K' = \mathbb{Q}(\sqrt{3})\). Of course, rational points of \(x^2 + y^2 = z^2\) are points of degree one.

Now let us take a look at curves over finite fields.

**Example 4.5.** Let \(C\) be a parabola over \(\mathbb{F}_2\) given by the equation \(yz = x^2\).

(a) First, it is easy to find all \(\mathbb{F}_2\)-points of \(C\):
\[
C(\mathbb{F}_2) = \{(0 : 0 : 1), (1 : 1 : 1), (0 : 1 : 0)\}.
\]

(b) Now consider the extension \(\mathbb{F}_2 \subset \mathbb{F}_4 = \{a + b\alpha \mid a, b \in \mathbb{F}_2\}\), where \(\alpha\) is a root of \(t^2 + t + 1 \in \mathbb{F}_2[t]\), and so \(\alpha^3 = 1\). We have
\[
C(\mathbb{F}_4) = \{(0 : 0 : 1), (1 : 1 : 1), (0 : 1 : 0), (\alpha : \alpha^2 : 1), (\alpha^2 : \alpha : 1)\}.
\]

Note that the first three elements of \(C(\mathbb{F}_4)\) are degree one points of \(C\) whereas the last two form a degree two point of \(C\). Indeed, it is easy to see that \(\alpha\) and \(\alpha^2\) are the two roots of \(t^2 + t + 1\). Hence the “conjugation”
\[
\sigma : \mathbb{F}_4 \to \mathbb{F}_4, \quad \sigma(a + b\alpha) = a + b\alpha^2
\]
together with the identity map \(\text{id} : \mathbb{F}_4 \to \mathbb{F}_4\) forms the Galois group \(\text{Gal}(\mathbb{F}_4/\mathbb{F}_2)\). If fact, \(\sigma\) is the Frobenius automorphism and, as we have already seen in Proposition 2.9 of Chapter 2, \(\text{Gal}(\mathbb{F}_4/\mathbb{F}_2)\) is cyclic generated by \(\sigma\).

(c) Enlarging the field even more, consider
\[
\mathbb{F}_2 \subset \mathbb{F}_8 = \{a + b\beta + c\beta^2 \mid a, b, c \in \mathbb{F}_2\},
\]
where \(\beta\) is a root of \(t^3 + t + 1 \in \mathbb{F}_2[t]\). Recall that \(\mathbb{F}_8\) is cyclic generated by \(\beta\), so \(\beta^7 = 1\) and
\[
\mathbb{F}_8 = \{0, 1, \beta, \ldots, \beta^6\}.
\]
As before the Frobenius automorphism \(\sigma\) which maps \(\beta\) to \(\beta^2\) generates the Galois group \(\text{Gal}(\mathbb{F}_8/\mathbb{F}_2)\):
\[
\text{Gal}(\mathbb{F}_8/\mathbb{F}_2) = \{\text{id}, \sigma, \sigma^2\}.
\]
Clearly, \(\sigma^2\) maps \(\beta\) to \(\beta^4\). We can list the elements of \(C(\mathbb{F}_8)\) according to the degree of the corresponding points of \(C\):
• three points of degree one:
  \((0 : 0 : 1), (1 : 1 : 1), (0 : 1 : 0)\)

• two points of degree three:
  \(\{(\beta : \beta^2 : 1), (\beta^2 : \beta^4 : 1), (\beta^4 : \beta : 1)\}, \)
  \(\{(\beta^3 : \beta^6 : 1), (\beta^6 : \beta^5 : 1), (\beta^5 : \beta^3 : 1)\}\)\

**Example 4.6.** Let \(C\) be the cubic given by \(y^2z = x^3 - xz^2 - z^3\) over \(F_3\).

(a) The affine part of \(C\) is given by \(y^2 = x^3 - x - 1\). Since \(y^2\) does not take value \(-1\) and \(x^3 = x\) for \(x, y \in F_3\) the affine curve has no \(F_3\)-points. We have
\[
C(F_3) = \{(0 : 1 : 0)\}.
\]

(b) Consider the degree two extension \(F_3 \subset F_9 = \{a + bi \mid a, b \in F_3\}\), where \(i\) is the root of \(t^2 = -1\). Here \(1 + i\) is a generator of \(F_9^*\) and we can write
\[
F_9 = \{0, \pm 1, \pm i, \pm (1 + i), \pm (1 - i)\}.
\]

It is easy to see that \(y^2\) takes only values \(0, \pm 1, \pm i\) in \(F_9\). Checking possible values of \(x\) in \(F_9\) we see that \(C\) has the following \(F_9\)-points:

• one point of degree one:
  \((0 : 1 : 0)\)

• three points of degree two:
  \(\{(0 : i : 1), (0 : -i : 1)\}, \)
  \(\{(1 : i : 1), (1 : -i : 1)\}, \)
  \(\{(-1 : i : 1), (-1 : -i : 1)\}\).

In [?] you can find the description of \(F_{27}\)-points of \(C\).

### 4.4. Hermitian Curves

**4.1.1. \(K\)-divisors and the Riemann–Roch formula.**

**4.2. The Zeta Function**

**4.3. The Hasse–Weil Bound**

**4.4. Hermitian Curves**
CHAPTER 5

Algebraic Geometry Codes

5.1. The \( \mathcal{L} \)-construction
5.2. The \( \Omega \)-construction
5.3. Duality
5.4. (Quasi-)Self-Dual AG codes
5.5. Asymptotics of AG codes