

Chapter II

Cubic Curves over Finite Fields

1. Rational points over finite fields

Our goal is to study cubic curves over a finite field, the field of integers modulo p . This means that the curve's polynomial equation will not have rational coefficients anymore but the coefficients will be in the finite field with p elements. We will denote this field with \mathbf{F}_p . We will study cubic equations $C : F(x, y) = 0$ with coefficients in the field \mathbf{F}_p and look for solutions (x, y) with $x, y \in \mathbf{F}_p$. More generally, we will look for solutions with $x, y \in \mathbf{F}_q$ where \mathbf{F}_q is an extension of \mathbf{F}_p with $q = p^e$ elements. We will call this solution point of curve C. If the coefficients x, y lie in \mathbf{F}_p , we will call it rational point of curve C .

Remark 2.1.1 Let a non-singular cubic curve defined over some field K . Then the sum of two rational points, here we consider points with coefficients in the field K , is defined exactly as in 1.5.4. With this law we construct the commutative group of the rational points of the curve.

Remark 2.1.2 Obviously, the remark 1.5.11 comes directly from the addition law so it holds for every field K .

Remark 2.1.3 The formulas 1.5.13, 1.5.14 and 1.5.15 for adding two rational points of an elliptic curve hold in every field K .

Consider the cubic curve

$$C : y^2 = x^3 + ax^2 + bx + c$$

for some $a, b, c \in \mathbf{F}_p$. We assume that $p \neq 2$. The curve is non-singular if and only if the discriminant $D = -4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2$ is not zero as an element of \mathbf{F}_q . (See, remark 1.4.9)

Working exactly as in the case that the polynomial $f(x)$ hasn't a 2-degree term, we compute the coefficients of the sum $P+Q$ of two rational points P, Q on C with $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ as follows:

$$x_3 = \lambda^2 - a - x_1 - x_2 \text{ and } y_3 = -(\lambda x_3 + v). \quad (2.1.4)$$

where

$$\lambda = \begin{cases} \frac{y_2 - y_1}{x_2 - x_1}, & \text{if } x_1 \neq x_2 \\ \frac{3x_1^2 + 2ax_1 + b}{2y_1}, & \text{if } P_1 = P_2 \end{cases}$$

and $v = y_1 - \lambda x_1 = y_2 - \lambda x_2$.

Notation 2.1.5 When the curve C is defined over a field K we will sometimes write $C|_K$.

Definition 2.1.6 Let $C|_K$. We denote the abelian group of its K -rational points with $C(K)$.

Before doing more general theory, let's look at an example. Consider the elliptic curve E given by the equation:

$$E: y^2 = x^3 + x + 1$$

defined over the field \mathbf{F}_5 . **How can we find the rational points?**

Since x and y are supposed to be in \mathbf{F}_5 , we can just take each of the five possibilities for x , apply them into the polynomial $x^3 + x + 1$ and check if the result is a quadratic residue in \mathbf{F}_5 .

We form the following table:

x	$x^3 + x + 1 = y^2$	quadratic residue	y
0	1	YES	1,-1
1	$1 + 1 + 1 = 3$	NO	-
2	$2^3 + 2 + 1 = 1$	YES	1,-1
3	$3^3 + 3 + 1 = 1$	YES	1,-1
4	$4^3 + 4 + 1 = 4$	YES	2,-2

So, including the point O "at infinity", we find nine points:

$$E(\mathbf{F}_5) = \{O, (0, \pm 1), (2, \pm 1), (3, \pm 1), (4, \pm 2)\}.$$

Thus, $E(\mathbf{F}_5)$ is an abelian group of order nine, it is isomorphic either with the cyclic group \mathbf{Z}_9 or with the $\mathbf{Z}_3 \times \mathbf{Z}_3$. We can determine which one by making the group table. Let $P = (0,1)$. Then using the formula 2.1.4 we find $2P = (x_3, y_3)$. (We keep in mind that we are working in \mathbf{F}_5).

We get:

$$a = 0, b = 1, x_1 = 0, y_1 = 0, \lambda = (3 \cdot 0^2 + 2 \cdot 0 \cdot 0 + 1) \cdot (2 \cdot 1)^{-1} = 2^{-1} = 3$$

$$v = y_1 - \lambda x_1 = 1 - 3 \cdot 0 = 1. \text{ So, } x_3 = 3^2 - 0 - 0 - 0 = 4 \text{ and } y_3 = -3 \cdot 4 - 1 = 2.$$

Thus, $2P = (4,2)$.

With the same way we find that $3P = (2,1)$ and $4P = (3, -1)$. So $\text{ord}(P) > 3$.

Hence, $E(\mathbf{F}_5)$ is a **cyclic group** of order nine. The point $P_1 = 3P = (2,1)$ has order 3. Order 3 has also the point $P_2 = -P_1 = (2, -1)$. All of the other non-zero points have order nine.

As these examples make clear because there are finite number of possibilities for x, y , the points (x, y) are finite and $C(\mathbf{F}_p)$ is a **finite group**. It is also clear that the order of $C(\mathbf{F}_p)$ is at maximum $2p+1$. A natural question that occurs is **how big is it?** Can we make any estimate as to the number of points in $C(\mathbf{F}_p)$?

2. Points of finite order

Now, we'll try to study a method for finding the points of finite order on a rational elliptic curve E with integer coefficients. The idea is to consider the curve locally. This means to reduce the curve for every prime p . So, if we consider the curve $E \bmod p$ by reducing its coefficients mod p we get a curve defined over the field \mathbf{F}_p . Of course the curve may have singularities. It will be an elliptic curve if and only if the discriminant D is not divisible by p .

Let C be a cubic curve, given by the equation:

$$C : y^2 = x^3 + ax^2 + bx + c$$

with integer coefficients a, b, c .

As we already know the group $C(\mathbf{Q})$ of rational points on curve C is finitely generated (Mordell's theorem) and the points of finite order have integer coordinates (Lutz – Nagell theorem).

We write $z \rightarrow \tilde{z}$ for the map “reduction modulo p ”,

$$Z \rightarrow \frac{Z}{pZ} = \mathbf{F}_p, z \mapsto \tilde{z}$$

Then we can take the equation the curve C , which has integer coefficients, and reduce those coefficients modulo p to get a new curve with coefficients in the field \mathbf{F}_p :

$$\tilde{C} : y^2 = x^3 + \tilde{a}x^2 + \tilde{b}x + \tilde{c}$$

When will the curve \tilde{C} be non-singular?

When $p \geq 3$ and the discriminant

$$\tilde{D} = -4\tilde{a}^3\tilde{c} + \tilde{a}^2\tilde{b}^2 + 18\tilde{a}\tilde{b}\tilde{c} - 4\tilde{b}^3 - 27\tilde{c}^2$$

is non-zero. But reduction modulo p is a homomorphism, so the discriminant \tilde{D} is just the reduction modulo p of the discriminant D of the cubic curve C . In other words, the curve \tilde{C} is non-singular if for the prime p holds $p \geq 3$ and $p \nmid D$.

Having reduced the curve C , it is natural to try making points on C and reducing them modulo p to get points on \tilde{C} . We can do this provided that the coordinates of the point have no p in their denominator. In particular, if a point has integer coordinates, then we can reduce that point modulo p for any prime p .

If $P = (x, y)$ is a rational point on C with integer coordinates, then x and y satisfy the equation:

$$y^2 = x^3 + ax^2 + bx + c.$$

This equation gives a relation among integers, so we can reduce it modulo p to get the equation:

$$\tilde{y}^2 = \tilde{x}^3 + \tilde{a}\tilde{x}^2 + \tilde{b}\tilde{x} + \tilde{c}$$

This tells us that $\tilde{P} = (\tilde{x}, \tilde{y})$ is a point in the group $\tilde{C}(\mathbf{F}_p)$. So we get a map from the points in $C(\mathbf{Q})$ with integer coordinates to $\tilde{C}(\mathbf{F}_p)$.

From the Lutz – Nagell theorem (theorem 1.5.9) we have that all points of finite order in $C(\mathbf{Q})$ have integer coordinates (and also that $y = 0$ or y divides the discriminant $D(f)$ of $f(x)$).

We now are going to study the set of points of finite order, which we'll denote with

$$\Phi := \{P = (x, y) \in C(\mathbf{Q}) : \text{ord}(P) < +\infty\}$$

Obviously, Φ is a **subgroup** of $C(\mathbf{Q})$ because if P_1, P_2 are points of finite order, say $m_1P_1 = O$ and $m_2P_2 = O$ then we get that $m_1m_2(P_1 - P_2) = m_1m_2P_1 - m_1m_2P_2 = O - O = O$. Hence $P_1 - P_2 \in \Phi$.

Since Φ consists of points with integer coefficients and O , we can define a reduction modulo p map

$$\left\{ \begin{array}{l} \Phi \rightarrow \tilde{C}(\mathbf{F}_p) \\ P \mapsto \tilde{P} = \begin{cases} (\tilde{x}, \tilde{y}) & \alpha v P = (x, y) \\ \tilde{O} & \alpha v P = O \end{cases} \end{array} \right.$$

Φ is a group, more precisely subgroup of $C(\mathbf{Q})$. If we choose $p \nmid 2D$ then we know that $\tilde{C}(\mathbf{F}_p)$ is also a group. So we have a map from the group Φ to the group $\tilde{C}(\mathbf{F}_p)$. We will prove that this map is a homomorphism of groups.

First we note that:

$$-P = (x, -y) = (\tilde{x}, -\tilde{y}) = -\tilde{P}$$

So it suffices to show that $P_1 + P_2 + P_3 = O \Rightarrow \tilde{P}_1 + \tilde{P}_2 + \tilde{P}_3 = O$. We have to consider various cases.

If any of P_1, P_2, P_3 is O , for example P_3 , we get: $P_1 + P_2 = O$ or $P_1 = -P_2$. Then, we have $\tilde{P}_1 = -\tilde{P}_2$ or $\tilde{P}_1 = -\tilde{P}_2$ or $\tilde{P}_1 + \tilde{P}_2 = \tilde{O}$, which is what we wanted to prove.

Let's assume that P_1, P_2, P_3 are not equal to O . We write their coordinates as:

$$P_1 = (x_1, y_1), \quad P_2 = (x_2, y_2), \quad P_3 = (x_3, y_3)$$

From the remark 2.1.2 the condition $P_1 + P_2 + P_3 = O$ is equivalent to say P_1, P_2, P_3 lie on a line. We have proven to the previous chapter that we can obtain the coordinates of P_3 from the relation 1.5.12

$$x^3 + ax^2 + bx + c - (\lambda x + v)^2 = (x - x_1)(x - x_2)(x - x_3).$$

This is the relation that ensures that $P_1 + P_2 + P_3 = O$, regardless of whether or not the points are distinct.

Reducing the last equation modulo p , we obtain:

$$x^3 + \tilde{a}x^2 + \tilde{b}x + \tilde{c} - (\tilde{\lambda}x + \tilde{v})^2 = (x - \tilde{x}_1)(x - \tilde{x}_2)(x - \tilde{x}_3)$$

Of course, we can also reduce the equations $y_i = \lambda x_i + v$ ($i \in \{1, 2, 3\}$) modulo p to get $\tilde{y}_i = \tilde{\lambda} \tilde{x}_i + \tilde{v}$ ($i \in \{1, 2, 3\}$). This means that the line $y = \tilde{\lambda}x + \tilde{v}$ intersects the curve at the three points $\tilde{P}_1, \tilde{P}_2, \tilde{P}_3$. Further, if two of the points $\tilde{P}_1, \tilde{P}_2, \tilde{P}_3$ are the same, say $\tilde{P}_1 = \tilde{P}_2$, then the line is the tangent of \tilde{C} at \tilde{P}_1 , and if $\tilde{P}_1 = \tilde{P}_2 = \tilde{P}_3$ then the line with the curve \tilde{C} have a triple order contact point. Therefore, in all cases:

$$\tilde{P}_1 + \tilde{P}_2 + \tilde{P}_3 = O.$$

This completes the proof that the reduction modulo p is a homomorphism from Φ to $\tilde{C}(\mathbf{F}_p)$.

Moreover, is a monomorphism (homomorphism and one-to-one) because a non-zero point (x, y) in Φ is sent to $(\tilde{x}, \tilde{y}) \in \tilde{C}(\mathbf{F}_p)$ which is not \tilde{O} . So the kernel of the reduction map modulo p consists only of O . This is equivalent to the fact that the map is one-to-one.

So, Φ is isomorphic to a subgroup of $\tilde{C}(\mathbf{F}_p)$ for every prime p ($p \nmid 2D$). This remark helps us in many cases to determine Φ with very little work.

Before giving some examples, we will restate formally the theorem we have just proved:

Theorem 2.2.1 (Reduction Modulo p Theorem) Let C be a non-singular curve

$$C : y^2 = x^3 + ax^2 + bx + c$$

with integer coefficients a, b, c and let D be the discriminant

$$D = -4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2.$$

Let $\Phi \subseteq C(\mathbf{Q})$ be the subgroup consisting of all points of C of finite order. For any prime p , let $P \rightarrow \tilde{P}$ be the reduction modulo p map:

$$\left\{ \begin{array}{l} \Phi \rightarrow \tilde{C}(\mathbb{F}_p) \\ P \mapsto \tilde{P} = \begin{cases} (\tilde{x}, \tilde{y}) & \text{if } P = (x, y) \\ \tilde{O} & \text{if } P = O \end{cases} \end{array} \right.$$

If $p \nmid 2D$ the reduction modulo p map is an isomorphism of Φ onto a **subgroup** of $\tilde{C}(\mathbb{F}_p)$.

Let's give three examples how to use this theorem to determine the points of finite order of a curve.

Example 2.2.2 $C : y^2 = x^3 + 3$

The discriminant for this curve is $D = -27 \cdot 3^2 = -243 = -3^5$, hence there is a monomorphism $\Phi \rightarrow \tilde{C}(\mathbb{F}_p)$ for all primes $p \geq 5$.

We will find the groups $\tilde{C}(\mathbb{F}_5)$ and $\tilde{C}(\mathbb{F}_7)$.

In the field \mathbb{F}_5 we have:

x	$x^3 + 3 = y^2$	quadratic residue	y
0	3	NO	-
1	$1+3 = 4$	YES	2, 3
2	$2^3+3 = 1$	YES	1, 4
3	$3^3+3 = 0$	YES	0
4	$4^3+3 = 2$	NO	-

So, including the point O "at infinity", we have that $\#\tilde{C}(\mathbb{F}_5) = 6$.

In the field \mathbb{F}_7 we have:

x	$x^3 + 3 = y^2$	quadratic residue	y
0	3	NO	-
1	4	YES	2, 5
2	4	YES	2, 5
3	2	YES	3, 4
4	4	YES	2, 5
5	2	YES	3, 4
6	2	YES	3, 4

So, including the point O "at infinity", we have that $\#\tilde{C}(\mathbb{F}_7) = 13$.

Because Φ is a subgroup of both $\tilde{C}(\mathbf{F}_5)$ and $\tilde{C}(\mathbf{F}_7)$ we have that $\#\Phi \mid \#\tilde{C}(\mathbf{F}_5)$ and $\#\Phi \mid \#\tilde{C}(\mathbf{F}_7)$. So $\#\Phi \mid 6$ and $\#\Phi \mid 13$. But $(6,13) = 1$. Thus, $\#\Phi = 1$. In other words, C has no points of finite order other than O .

We note that $(1, 2) \in C(\mathbf{Q})$ and, according to what we have proven, it has infinite order. So the curve C has infinitely many rational points.

It is worth comparing this method for determining Φ with the procedure given by the Lutz-Nagell theorem (theorem 1.5.9) and the proposition 1.5.10. Using the proposition 1.5.10 we have to prove that they are no points on C such that the square of their y coordinate to divide -243 or in other words, with y coordinate such that $y \in \{\pm 1, \pm 3, \pm 9, \pm 27, \pm 81\}$. Clearly if $y = \pm 1$ we have $1 = x^3 + 3x$ which gives us no rational points. Let now $3 \mid y$ then the equation $y^2 = x^3 + 3x$ gives us that $3 \mid x$ but also we can write $3 = y^2 - x^3$. So because $9 \mid y^2 - x^3$ we have that $9 \mid 3$ which is a contradiction. Thus, with the Lutz – Nagell theorem, we have proven that $\#\Phi = 1$.

Example 2.2.3

$$C: y^2 = x^3 + x$$

The discriminant for this curve is $D = -4 \cdot 1^3 = -4$. Because the discriminant is quite small it might be easiest to use the Lutz – Nagell theorem but we will use the reduction theorem. We have a one-to-one map $\Phi \rightarrow \tilde{C}(\mathbf{F}_p)$ for every prime $p \geq 3$.

We do some little computation:

In the field \mathbf{F}_3 we have:

x	$x^3 + x = y^2$	y
0	0	0
1	2	-
2	1	1, 2

So, including the point O “at infinity”, we have that $\#\tilde{C}(\mathbf{F}_3) = 4$.

In the field \mathbf{F}_5 we have:

x	$x^3 + x = y^2$	y
0	0	0
1	2	-
2	0	0
3	0	0
4	3	-

So, including the point O “at infinity”, we have that $\#\tilde{C}(\mathbf{F}_5) = 4$.

In the field \mathbf{F}_7 we have:

x	$x^3 + x = y^2$	y
---	-----------------	---

0	0	0
1	2	3, 4
2	3	-
3	2	3, 4
4	5	-
5	4	2, 5
6	5	-

So, including the point O “at infinity”, we have that $\#\tilde{C}(\mathbf{F}_7) = 8$.

We will prove that 4 divides $\#\tilde{C}(\mathbf{F}_p)$ for all primes $p \geq 3$.

Obviously, two points on the curve

$$\tilde{C} : y^2 = x(x^2 + 1) \quad (1)$$

are O and $(0, 0)$.

We now check two cases.

- If $p \equiv 1 \pmod{4}$ then $\left(\frac{-1}{p}\right) = 1$ so there exists an $x_0 \in \mathbf{F}_p$ such that $x_0^2 \equiv -1 \pmod{p}$.

Thus, the points $(x_0, 0)$ and $(-x_0, 0)$ are also points of (1). Until now, we have found four points.

Moreover, if there is some x_1 such that $x_1(x_1^2 + 1) = c$ where c is a non-zero quadratic residue mod p then the equation $y^2 = c$ has two solutions, let y_1 and y_2 . This means that the points (x_1, y_1) and (x_1, y_2) are points of (1). Also we have that

$-x_1((-x_1)^2 + 1) = -c$ and $\left(\frac{-c}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{c}{p}\right) = \left(\frac{c}{p}\right)$. So, the equation $y^2 = -c$ has

also two solutions, let y_3 and y_4 , and the points $(-x_1, y_3)$ and $(-x_1, y_4)$ are also in (1). So, we take pairs non-zero quadratic residues and each one gives us two points.

So, $4 \mid \#\tilde{C}(\mathbf{F}_p)$.

- If $p \equiv 3 \pmod{4}$ we have that $\left(\frac{-1}{p}\right) = -1$ and then it holds that

$$\left(\frac{-c}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{c}{p}\right) = -\left(\frac{c}{p}\right).$$

If there is any x_1 such that $x_1(x_1^2 + 1) = c$ where c is a non-zero quadratic residue mod p then the equation $y^2 = c$ has two solutions. If c is not a quadratic residue mod p then for $-x_1$ we have that $-x_1((-x_1)^2 + 1) = -c$ where $-c$ is a quadratic residue mod

p. So, including the two obvious solutions, we have totally $\frac{p-1}{2} \cdot 2 + 2 = p + 1$ points on \tilde{C} . It holds that $p + 1 \equiv 3 + 1 \equiv 0 \pmod{4}$. So, again, $4 \mid \#\tilde{C}(\mathbf{F}_p)$.

Let's examine the groups $\tilde{C}(\mathbf{F}_3)$ and $\tilde{C}(\mathbf{F}_5)$:

$$\begin{aligned}\tilde{C}(\mathbf{F}_3) &= \{O, (0, 0), (2, 1), (2, 2)\} \\ \tilde{C}(\mathbf{F}_5) &= \{O, (0, 0), (2, 0), (3, 0)\}\end{aligned}$$

Because for any point $P = (x, y)$ it holds $-P = (x, -y)$ we have that a point P has order two if and only if $y = 0$. So $\tilde{C}(\mathbf{F}_3)$ has only one point of order two while $\tilde{C}(\mathbf{F}_5)$ has three. So

$$\tilde{C}(\mathbf{F}_3) \cong \mathbf{Z}_4, \quad \tilde{C}(\mathbf{F}_5) \cong \mathbf{Z}_2 \times \mathbf{Z}_2$$

Because $\Phi \leq \tilde{C}(\mathbf{F}_3)$ and $\Phi \leq \tilde{C}(\mathbf{F}_5)$ the only possibilities are that Φ is trivial or cyclic of order two. We notice that $(0, 0) \in C(\mathbf{Q})$ is a point of order two, we conclude that $\Phi = \{O, (0, 0)\}$.

Example 2.2.4

$$C: y^2 = x^3 - 43x + 166$$

The discriminant for this curve is $D = -4(-43)^3 - 27 \cdot 166^2 = -425984 = -2^{15} \cdot 13$.

We will try a point with integer coefficients using the stronger form of the Lutz – Nagell theorem (proposition 1.5.10). Let $P = (x, y) \in C$ a point with integer coefficients. It holds that $y^2 \mid D$.

We try for $y = \pm 1$:

We have $1 = x^3 - 43x + 166$ i.e. $x(x^2 - 43) = -165$ i.e. $x(x^2 - 43) = -3 \cdot 5 \cdot 11$. Because the equation $x^2 \equiv 43 \pmod{p}$ has no solutions for $p = 3, 5, 11$ it must be $x^2 - 43 = \pm 1$. So, $x^2 = 42$ or $x^2 = 44$. This is a contradiction for every integer x .

We try for $y = \pm 2$:

We have $4 = x^3 - 43x + 166$ i.e. $x(x^2 - 43) = -162$ i.e. $x(x^2 - 43) = -2 \cdot 3^4$. Because the equation $x^2 \equiv 43 \pmod{3}$ has no solutions it must be $x^2 - 43 = \pm 1, \pm 2$. So $x^2 = 41, 42, 44, 45$. This is a contradiction for every integer x .

We try for $y = \pm 4$:

We have $16 = x^3 - 43x + 166$ i.e. $x(x^2 - 43) = -150$ i.e. $x(x^2 - 43) = -2 \cdot 3 \cdot 5^2$. Because the equation $x^2 \equiv 43 \pmod{p}$ has no solutions for $p = 3, 5$ it must be $x^2 - 43 = \pm 1, \pm 2$. So $x^2 = 41, 42, 44, 45$. This is a contradiction for every integer x .

We try for $y = \pm 8$:

We have $64 = x^3 - 43x + 166$ i.e. $x(x^2 - 43) = -102$. This equation has solution for $x = 3$.

So the point $P = (3, 8)$ is of finite order.

To determine the order of the element P let's take the reduction of curve C modulo 5 and calculate the group $\tilde{C}(\mathbf{F}_5)$:

$$\tilde{C} : y^2 = x^3 + 2x + 1$$

x	$x^3 + 2x + 1 = y^2$	y
0	1	1, 4
1	4	2, 3
2	3	-
3	4	2, 3
4	3	-

So, including the point O "at infinity", we have that $\#\tilde{C}(\mathbf{F}_5) = 7$.

Hence, if $\text{ord}(P) > 7$ then $\text{ord}(P) = +\infty$. We calculate $2P$:

$$a = 0, b = -43, x_1 = 3, y_1 = 8, \lambda = \frac{3 \cdot 3^2 - 43}{2 \cdot 8} = -1, v = y_1 - \lambda x_1 = 8 + 3 = 11.$$

So, $2P = (x_3, y_3)$ where $x_3 = (-1)^2 - 6 = -5$ and $y_3 = -5 - 11 = -16$. Thus, we have that $2P = (-5, -16)$. Similarly we find that $3P = (11, -32)$, $4P = (11, 32)$, $5P = (-5, 16)$ and $6P = (3, -8) = -P$. So $7P = O$ and $\text{ord}(P) = 7$. Because $\Phi \leq \tilde{C}(\mathbf{F}_5)$ and $P \in \Phi$ it holds that $\Phi \cong \mathbf{Z}_7$ and more precisely

$$\Phi = \langle P \rangle = \{O, (3, \pm 8), (-5, \pm 16), (11, \pm 32)\}.$$

A difficult problem is to determine which values are likely for the order of points of finite order on an elliptic curve. We report without proof very beautiful but also very difficult

Theorem 2.2.5 (Mazur, 1976) Let E be a non-singular cubic curve defined over \mathbf{Q} . Let $E(\mathbf{Q})$ has a point of finite order m . Then $1 \leq m \leq 10$ or $m = 12$. In particular, the group of the points of finite order of E is isomorphic with one of the following groups:

- (i) \mathbf{Z}_N with $1 \leq N \leq 10$ or $N = 12$
- (ii) $\mathbf{Z}_2 \times \mathbf{Z}_{2N}$ with $1 \leq N \leq 4$.

(See [S1], page 58)

3. The Riemann hypothesis

(for algebraic function field of one variable with coefficients from the finite field \mathbf{F}_q)

We have seen for each prime p , there is a field \mathbf{F}_p of p elements. In fact, given any prime p and an integer $r \geq 1$, there is exactly one field \mathbf{F}_q of $q = p^r$ elements. The field \mathbf{F}_q has \mathbf{F}_p as a subfield and for each α in \mathbf{F}_q it holds $p\alpha = 0$. Conversely, any finite field is isomorphic to a field \mathbf{F}_q for some $q = p^r$. The field \mathbf{F}_q is characterized by the property all its elements to be exactly the roots of the polynomial

$$f(X) = X^q - X$$

this means

$$f(X) = \prod_{\alpha \in \mathbf{F}_q} (X - \alpha)$$

Proposition 2.3.1 Let K be a field which contains \mathbf{F}_q . The map $\text{Fr}: K \rightarrow K$ defined by $\text{Fr}(x) = x^q$ is an \mathbf{F}_q -**endomorphism** of the ring K . (Endomorphism with the property $\text{Fr}(\alpha) = \alpha$ for every $\alpha \in \mathbf{F}_q$).

Proof Let $x, y \in K$. Then

1. $(x + y)^q = x^q + y^q$
2. $(xy)^q = x^q y^q$ and $\alpha^q = \alpha$ if $\alpha \in \mathbf{F}_q$.

The second assertion needs no proof.

The first follows from the fact that for the binomial coefficients

$$(x + y)^q = \sum_{j=0}^q \binom{q}{j} x^j y^{q-j}$$

for $j = 1, \dots, q - 1$, it holds $\binom{q}{j} = \frac{q!}{j!(q-j)!} = \frac{(q-j+1)(q-j+2) \cdots q}{1 \cdot 2 \cdots k}$. We can write

the last equality in the form $(q-j+1)(q-j+2) \cdots q = \binom{q}{j} \cdot 1 \cdot 2 \cdots k$. We have

that $q = p^r$ divides the left part of the equality and does not divide $1, 2, \dots, k$ so it divides $\binom{q}{j}$, and because the field characteristic is p , only the first and the last terms

survive., because $\binom{q}{0} = \binom{q}{q} = 1$ and all the other terms are equal to zero.

So the map $\text{Fr}: K \rightarrow K$ is an \mathbf{F}_q -endomorphism of the ring K . It is called the **Frobenius endomorphism**. The Frobenius endomorphism is physiologically extended, at components, in the affine and the projective space.

We are interesting in counting the number N_q of solutions in $\mathbf{F}_q \times \mathbf{F}_q$ of the equation

$$Y^2 = f(X)$$

where $f(X) = AX^3 + BX^2 + CX + D \in \mathbf{F}_q[X]$, is a polynomial of the third degree ($A \neq 0$) without repeated roots (it has only single roots). We suppose $p \neq 2, 3$ so, as we have show in chapter 1, the equation can be written in the Weierstrass form (see definition 1.4.7):

$$Y^2 = X^3 + bX + c$$

for some $b, c \in \mathbf{F}_q$. Together with the point O “at infinity”, these solutions form an abelian group of order $N'_q = N_q + 1$. This is the group of the \mathbf{F}_q -rational points on the elliptic curves E defined by the last equation. Let $q = p$. In 1924, Artin conjectured the following estimate for N_p : $|N_p - p| \leq 2\sqrt{p}$. Actually, the equivalent form of this inequality is the analog for the field of rational functions on the curve E of what Riemann conjectured much earlier for the field of rational numbers, and is well known as the Riemann hypothesis. Gauss was the first to study the behavior of N_p for the different values of p , for the curve

$$Y^2 = X^3 - 432$$

In fact, he gave a precise formula for N_p .

Theorem 2.3.2 (Gauss, 1801) Let N_p be the number of solutions in $\mathbf{F}_p \times \mathbf{F}_p$ of the equation $Y^2 = X^3 - 432$, $p \neq 2, 3$. Then

1. $N_p = p$ for $p \equiv 2 \pmod{3}$
2. If $p \equiv 1 \pmod{3}$, there are integers A, B unique up to sign, such that $4p = A^2 + 27B^2$. If the sign of A is chosen that $A \equiv 1 \pmod{3}$, then $N_p = p + A - 2$. In particular, $|N_p - p| \leq 2\sqrt{p}$.

Artin’s conjecture was proved by Hasse in 1936. Later in 1948 Weil generalized it to his famous theorem (the Riemann hypothesis for curves over finite fields) and made some intriguing conjectures, which are known as Weil conjectures.

Theorem 2.3.3 The Riemann hypothesis for curves over finite Fields (Weil). The number N_q of points with coordinates in \mathbf{F}_q on an irreducible, non-singular curve defined over \mathbf{F}_q and of genus g satisfies

$$|N_q - q| \leq 2g\sqrt{q} \quad (1)$$

Manin gave a completely elementary proof of Hasse’s theorem and a valuation theoretic proof is due to Zimmer. Weil’s proof of Riemann hypothesis depends heavily in algebraic geometry. A somewhat simpler proof was given by Roquette. An

elementary proof was later initiated by Stepanov and completed by W. Schmidt. A very elegant but less elementary proof based on Stepanov's method is by Bombieri. More about the history of the Riemann hypothesis can be found in [Ch1] and [Ch2]. We shall give Manin's proof of Hasse's theorem.

Theorem 2.3.4 (Hasse, 1936) Let $p \neq 2, 3$. The number N_p of solutions in $\mathbf{F}_p \times \mathbf{F}_p$ of the elliptic curve

$$Y^2 = X^3 + bX + c$$

where $a, b \in \mathbf{F}_p$ with $\Delta = -4b^3 - 27c^2$ in \mathbf{F}_p^* satisfies the inequality

$$|N_p - p| \leq 2\sqrt{p}$$

Remark 2.3.5 When the curve is projective, there is an extra point (the point at infinity), so that the total number of points is $N'_q = N_q + 1$ and the last inequality becomes

$$|N'_q - (q + 1)| \leq 2g\sqrt{q}$$

Now we are going to explain why Hasse's theorem is called the Riemann hypothesis for the elliptic curve. First, recall that the Riemann ζ -eta function is defined as $\zeta(s) = \sum_{n \in \mathbf{N}} n^{-s}$, for $s \in \mathbf{C}$ with $\text{Re}(s) > 1$. It is proved that the function converges uniformly in all the complex plane, except the simple pole for $s = 1$. Moreover it verifies a functional equation which connects $\zeta(s)$ with $\zeta(1 - s)$. To be more specific, if $\xi(s) = \pi^{-\frac{s}{2}} \Gamma(\frac{s}{2}) \zeta(s)$ then $\xi(s) = \xi(1 - s)$, where $\Gamma(s)$ is the well-known Γ -function.

Riemann hypothesis is reported in the conjecture that all the roots of $\zeta(s)$ for $0 \leq \text{Re}(s) \leq 1$ lay on the line $\text{Re}(s) = \frac{1}{2}$. Let E be an elliptic curve over a finite field \mathbf{F}_p . The zeta function for that curve is defined as follows

$$Z(E|_{\mathbf{F}_p}, T) = \frac{1 - \alpha_p T + pT^2}{(1 - T)(1 - pT)},$$

where $\alpha_p = \alpha_p(E) = p - N_p$. We omit here to discuss properties of that function equivalent to properties of the Riemann ζ -eta function. Due to the theorem of Hasse, the polynomial $pT^2 - \alpha_p T + 1$ has as discriminant

$$\Delta(E) = \alpha_p^2 - 4p < 0.$$

The two roots of the polynomial are complex conjugates, let a and \bar{a} . So we get

$$1 - \alpha_p T + pT^2 = (1 - aT)(1 - \bar{a}T)$$

where $a + \bar{a} = \alpha_p$ and $|a| = |\bar{a}| = \sqrt{p}$.

If we consider $T = p^{-s}$ where $s = \sigma + it \in \mathbf{C}$ and define

$$\zeta(E|_{\mathbb{F}_p}, s) := Z(E|_{\mathbb{F}_p}, p^{-s}) = \frac{1 - \alpha_p p^{-s} + p^{1-2s}}{(1 - p^{-s})(1 - p^{1-s})}$$

we get $\zeta(E|_{\mathbb{F}_p}, s) = 0$ if and only if $1 - \alpha_p p^{-s} = 0$ or $1 - \bar{a} p^{-s} = 0$. Thus, $p^s = a$ or $p^s = \bar{a}$.

Now, because $|p^s| = p^{\operatorname{Re}(s)} = p^\sigma$ we have that $|p^s| = \sqrt{p}$ is equivalent with $\sigma = \frac{1}{2}$, the truth for the Riemann hypothesis for $\zeta(E|_{\mathbb{F}_p}, s)$. We note that the Riemann hypothesis for the function $\zeta(E|_{\mathbb{F}_p}, s)$ is equivalent with the fact that $|a| = |\bar{a}| = \sqrt{p}$. For the polynomial $P(T) = 1 - \alpha_p T + pT^2$ we have that $|\alpha_p| = |a + \bar{a}| \leq |a| + |\bar{a}| = 2\sqrt{p}$. Thus, $|N_p - p| \leq 2\sqrt{p}$. We just proved that the Riemann hypothesis for the function $\zeta(E|_{\mathbb{F}_p}, s)$ is equivalent to Hasse's theorem.

4. Manin's proof of the Hasse inequality

First let us remark that Hasse's theorem:

If p prime different than 2 and 3 and N_p the number of solutions in $F_p \times F_p$ of the equation

$$Y^2 = X^3 + bX + c$$

where $b, c \in F_q$ with $\Delta = -4b^3 - 27c^2$ in F_p^* then N_p satisfies the inequality

$$|N_p - p| \leq 2\sqrt{p},$$

holds and in the case we replace p with $q = p^r$.

Definition 2.4.1 Let E_1, E_2 be two elliptic curves defined over a field K . Then E_1 is a **twist** of E_2 if E_1 and E_2 become isomorphic over a finite extension L of K .

Example 2.4.2 Let E_1, E_2 be two elliptic curves over the field \mathbf{Q} of the rational numbers, which are given by the equations

$$E_1 : Y^2 = X^3 + bX + c$$

$$E_2 : d Y^2 = X^3 + bX + c$$

where $a, b, d \in \mathbf{Z}$ and d is square-free.

E_1 and E_2 are not isomorphic over \mathbf{Q} , but if we consider the finite extension $L = \mathbf{Q}(\sqrt{d})$, then E_2 is written over L :

$$E_2 : (\sqrt{d} Y)^2 = X^3 + bX + c$$

Now there is an isomorphism $\Phi : E_1 \rightarrow E_2$ which is defined by $(X, Y) \mapsto (X, \sqrt{d} Y)$. So $E_1|_{\mathbf{Q}}$ is a twist of $E_2|_{\mathbf{Q}}$.

Proposition 2.4.3 Let $K = \mathbf{F}_p(x)$ be the function field in one variable over \mathbf{F}_p and suppose E is the elliptic curve defined by the equation:

$$E : Y^2 = X^3 + bX + c \quad (b, c \in \mathbf{F}_p) \quad (1)$$

E is defined over \mathbf{F}_p so is also defined over $K = \mathbf{F}_p(x)$

If E_λ is the elliptic curve defined by the equation

$$E_\lambda : \lambda Y^2 = X^3 + bX + c \quad (2)$$

where $\lambda = \lambda(x) = x^3 + bx + c \in K = \mathbf{F}_p(x)$

then E and E_λ are both defined over K and E_λ is a twist of E over the field K .

Proof Like the example, we consider the finite extension $L = K(\sqrt{\lambda})$, then E and E_λ are defined over L and E_λ is written:

$$E_\lambda : (\sqrt{\lambda} Y)^2 = X^3 + bX + c$$

Then **there is** an isomorphism $\Phi_\lambda : E \rightarrow E_\lambda$ given by $(X, Y) \mapsto (X, \sqrt{\lambda} Y)$. So $E|_K$ is a twist of $E_\lambda|_K$.

To prove Hasse's theorem we shall consider the group $E_\lambda(K)$ of K -rational points on E_λ .

Let $P = (X_1, Y_1)$, $Q = (X_2, Y_2)$, $P+Q = (X_3, Y_3) = (X_1, Y_1) + (X_2, Y_2) \in E_\lambda(K)$. We will give some formulas similar to 1.5.14 and 1.5.15 that will allow us to compute $P+Q$ easily.

If $X_1 \neq X_2$, exactly as in chapter 1, we write the line L which is defined by P and Q in the form $Y = \kappa X + v$ where κ and v are given by the formulas $\kappa = \frac{Y_2 - Y_1}{X_2 - X_1}$ and

$v = Y_1 - \kappa X_1 = Y_2 - \kappa X_2$. Because of its construction L intersects the curve E on points P , Q . To calculate the third intersection point PQ we substitute Y from the last relation in the curve's equation.

$$\lambda(\kappa X + v)^2 = X^3 + bX + c$$

or else

$$X^3 - \lambda\kappa^2 X^2 + (b - 2\lambda\kappa v)X + (c - v^2) = 0$$

The last relation is a third degree equation in X and has roots X_1, X_2, X_3 so:

$$X^3 - \lambda\kappa^2 X^2 + (b - 2\lambda\kappa v)X + (c - v^2) = (X - X_1)(X - X_2)(X - X_3) \quad (2.4.4)$$

Hence, we get: $X_1 + X_2 + X_3 = -(-\lambda\kappa^2)$ or

$$X_3 = \lambda\kappa^2 - X_1 - X_2 \text{ and } Y_3 = \kappa X_3 + v \quad (2.4.5)$$

so, by substituting κ and v we get

$$X_3 = \lambda \left(\frac{Y_2 - Y_1}{X_2 - X_1} \right)^2 - (X_1 + X_2) \text{ and } Y_3 = \frac{Y_2 - Y_1}{X_2 - X_1} (X_3 - X_1) + Y_1 \quad (2.4.6)$$

Similarly we work if $(X, Y) = 2(X_1, Y_1)$

From the relation $\lambda Y^2 = X^3 + bX + c$ we get that $\kappa = \left. \frac{dY}{dX} \right|_{(X_1, Y_1)} = \frac{3X_1^2 + b}{2\lambda Y_1}$. By substituting κ from the formulas 1.5.12 and λY^2 by $X^3 + aX + b$ we find that:

$$X_3 = \frac{(3X_1^2 + b)^2}{4(X_1^3 + bX_1 + c)} - 2X_1 = \frac{X_1^4 - 2bX_1^2 - 8cX_1 + b^2}{4(X_1^3 + bX_1 + c)} \quad (2.4.7)$$

If in the equation (2) we substitute $X = x \in K = \mathbb{F}_p(x)$, we get:

$$\lambda Y^2 = x^3 + bx + c \quad \text{and} \quad \lambda Y^2 = \lambda \eta Y^2 = 1 \quad \eta Y = \pm 1.$$

So, two obvious solutions of (2) are

$$(x, 1) \text{ and } (x, -1) = -(x, 1)$$

A less obvious solution is:

$$X_0 = x^p, \quad Y_0 = (x^3 + bx + c)^{\frac{p-1}{2}}$$

That's true because if (X_0, Y_0) is a point of E_λ with $X_0 = x^p$ then $\lambda Y_0^2 = (x^p)^3 + bx^p + c$ or due to proposition 2.3.1:

$$\lambda Y_0^2 = (x^3 + bx + c)^p$$

It holds that: $(x^3 + bx + c)^p = (x^3 + bx + c) (x^3 + bx + c)^{p-1} = \lambda \left((x^3 + bx + c)^{\frac{p-1}{2}} \right)^2$.

Thus, $\lambda Y_0^2 = \lambda \left((x^3 + bx + c)^{\frac{p-1}{2}} \right)^2 \quad \text{and} \quad Y_0^2 = \left((x^3 + bx + c)^{\frac{p-1}{2}} \right)^2$.

This last equality completes the proof because

$$(x^3 + bx + c)^{\frac{p-1}{2}} = \lambda^{\frac{p-1}{2}} \in K.$$

$$\text{Let } (X_n, Y_n) = (X_0, Y_0) + n(x, 1) \quad n \in \mathbf{Z} \quad (2.4.8)$$

If $(X_n, Y_n) \neq 0$ we can prove that $X_n \neq 0$. (See lemma 2.5.1 on the next paragraph)

Writing X_n in the lowest form as $X_n = \frac{P_n}{Q_n}$ where $Q_n, P_n \in \mathbb{F}_p[x]$ and P_n is monic, we

get a well-defined function:

$$d: \mathbf{Z} \rightarrow \{0, 1, 2, 3, \dots\}$$

given by

$$d(n) = d_n = \begin{cases} 0, & \text{if } (X_n, Y_n) = 0 \\ \deg(P_n), & \text{otherwise} \end{cases}$$

Due to the definition of P_n the function d is well defined.

Manin's proof of Hesses's theorem is based on the following **basic identity**:

$$\text{BASIC IDENTITY: } d_{n-1} + d_{n+1} = 2d_n + 2.$$

(The proof of the basic identity can be found in paragraph 2.5)

The connection between Hasse's theorem and the function $d(n)$ is the following identity:

$$d_{-1} - d_0 - 1 = N_p - p \quad (2.4.9)$$

To prove identity (2.4.9) we need to put the rational function X_{-1} in the lowest form.

By formula (2.4.8) we have that $(X_{-1}, Y_{-1}) = (X_0, Y_0) + (x, -1)$. By the addition formula (2.4.6) we get:

$$\begin{aligned} X_{-1} &= \lambda \frac{\left[(x^3 + bx + c)^{\frac{p-1}{2}} + 1 \right]^2}{(x^p - x)^2} - (x^p + x) = \\ &= \frac{(x^3 + bx + c) \left[(x^3 + bx + c)^{\frac{p-1}{2}} + 1 \right]^2}{(x^p - x)^2} - (x^p + x) = \\ &= \frac{\lambda \left(\lambda^{\frac{p-1}{2}} + 1 \right)^2 - (x^p + x)(x^p - x)^2}{(x^p - x)^2} = \\ &= \frac{\lambda \left(\lambda^{p-1} + 2\lambda^{\frac{p-1}{2}} + 1 \right) - (x^p + x)(x^p - x)^2}{(x^p - x)^2} = \\ &= \frac{\left(\lambda^p + 2\lambda^{\frac{p+1}{2}} + \lambda \right) - (x^p + x)(x^p - x)^2}{(x^p - x)^2} = \\ &= \frac{\left((x^3 + bx + c)^p + 2\lambda^{\frac{p+1}{2}} + \lambda \right) - (x^{3p} - x^{2p+1} - x^{p+2} + x^3)}{(x^p - x)^2}. \end{aligned}$$

So using the proposition 2.3.1

$$\begin{aligned}
X_{-1} &= \frac{\left((x^{3p} + bx^p + c) + 2\lambda \frac{p+1}{2} + \lambda \right) - (x^{3p} - x^{2p+1} - x^{p+2} + x^3)}{(x^p - x)^2} = \\
&= \frac{x^{3p} + bx^p + c + 2\lambda \frac{p+1}{2} + \lambda - x^{3p} + x^{2p+1} + x^{p+2} - x^3}{(x^p - x)^2} = \\
&= \frac{x^{2p+1} + x^{p+2} + bx^p + 2\lambda \frac{p+1}{2} + \lambda - x^3 + c}{(x^p - x)^2}.
\end{aligned}$$

λ is a third degree polynomial in x and the polynomial $\lambda \frac{p+1}{2}$ is of $3 \frac{p+1}{2}$ degree which is less than $2p$. So,

$$X_{-1} = \frac{x^{2p+1} + R(x)}{(x^p - x)^2}$$

where $R(x)$ is a polynomial of degree at most $2p$.

To put X_{-1} in the **lowest** form $\frac{P_{-1}}{Q_{-1}}$ we first note that

$$(x^p - x) = x(x-1) \dots (x-p+1)$$

So the denominator is written:

$$(x^p - x)^2 = x^2 (x-1)^2 \dots (x-p+1)^2$$

Hence, as we can see from the first equality we gave for X_{-1} (see above) the only factors to cancel from the denominator are these who divide the factors in the

nominator $\left((x^3 + bx + c) \frac{p-1}{2} \right)^2 + 1$ or $\lambda = x^3 + bx + c$. So, these are:

- either $(x-r)^2$ when $(r^3 + rx + c) \frac{p-1}{2} = -1$ which that means that to the Legendre symbols it holds $\left(\frac{r^3 + br + c}{p} \right) = -1$
- either $x-r$ when $r^3 + br + c = 0$ ($0 \leq r < p$).

If m is the number of factors of the first kind and n the number of factors of the second kind, then

$$d_{-1} = \deg P_{-1} = 2p + 1 - 2m - n$$

But,

$$d_0 = \deg(P_0) = \deg(x^p) = p \quad (2.4.10)$$

So that

$$d_{-1} - d_0 = p + 1 - 2m - n \quad (2.4.11)$$

We also note that each r in \mathbb{F}_p with $r^3 + ar + b \neq 0$ and $\left(\frac{r^3 + br + c}{p}\right) = 1$ will give us two solutions and we will get only one solution from $r^3 + br + c = 0$. We have that $\left(\frac{r^3 + br + c}{p}\right) = 1$ it holds for $p - m - n$ values of r whereas $r^3 + br + c = 0$ holds for n values.

Hence, $N_p = 2(p - m - n) + n \hat{=} N_p = 2(p - m) - n$ and (2.4.9) follows from (2.4.11).

Lemma 2.4.12 The function $d(n)$ is a quadratic polynomial in n . In fact,

$$d_n = n^2 - (d_{-1} - d_0 - 1)n + d_0$$

Proof

Firstly we prove the lemma for $n = -1$ and $n = 0$.

For $n = -1$ the equality is written

$$d_{-1} = (-1)^2 - (d_{-1} - d_0 - 1)(-1) + d_0 \hat{=} d_{-1} = 1 + d_{-1} - d_0 - 1 + d_0$$

So lemma is true for $n = -1$.

For $n = 0$ the equality is written

$$d_0 = 0^2 - (d_{-1} - d_0 - 1)0 + d_0 \hat{=} d_0 = d_0.$$

So lemma is true for $n = 0$.

We continue using the mathematical induction in n .

Suppose it is true for $n - 1$ and n ($n \geq 0$).

By the basic identity,

$$\begin{aligned} d_{n+1} &= 2d_n - d_{n-1} + 2 \\ &= 2[n^2 - (d_{-1} - d_0 - 1)n + d_0] - [(n-1)^2 - (d_{-1} - d_0 - 1)(n-1) + d_0] + 2 \\ &= (n+1)^2 - (d_{-1} - d_0 - 1)(n+1) + d_0. \end{aligned}$$

This proves the lemma for $n + 1$. By induction the lemma holds for all $n \geq -1$. Similarly, it holds for all $n \leq 0$.

Proof of Hasse's Theorem

We define the quadratic polynomial

$$d(x) = x^2 - (d_{-1} - d_0 - 1)x + d_0 = x^2 - (N_p - p)x + d_0$$

We have shown $d_0 = p$ (see 2.4.10). So the discriminant of the polynomial $d(x)$ is

$$D = (N_p - p)^2 - 4p.$$

If we show that $D \leq 0$ for any prime p we have the proof of Hasse's theorem. The discriminant D cannot be positive, for any prime p , because otherwise the polynomial $d(x)$ will have two distinct real roots. Let α, β ($\alpha < \beta$) be these roots. Between the two roots the polynomial gives only negative values. Because, for every $n \in \mathbf{Z}$ it holds that $d(n) \geq 0$, we get that the two roots of the polynomial will be between two successive integers this means that there exists a $n_0 \in \mathbf{Z}$ such that:

$$n_0 \leq \alpha < \beta \leq n_0 + 1$$

Moreover, both equalities cannot hold simultaneously because by definition, $d(n)$ cannot be zero for two successive integers because if we had $d_{n_0} = 0$ then $(X_{n_0}, Y_{n_0}) = 0$ but then $(X_{n_0+1}, Y_{n_0+1}) = (X_{n_0}, Y_{n_0}) + (x, 1) = (x, 1)$ and $d(n_0 + 1) = 1 \neq 0$.

So from α, β at most only one can be an integer. Due to the inequality we get that $(\alpha - \beta) \notin \mathbf{Z}$. But this is a **contradiction** because α, β are roots of $d(x)$ and that means $D = (\alpha + \beta)^2 - 4\alpha\beta = (\alpha - \beta)^2$. Thus, it should $D \in \mathbf{Z}$ and simultaneously $(\alpha - \beta)^2 \notin \mathbf{Z}$

So $(N_p - p)^2 - 4p \leq 0$ which proves the theorem.

Manin's proof is included inter alia, and in the book of Franz Lemmermeyer, *Elliptische Kurven I*, (Available at World Wide Web: <http://www.rzuser.uni-heidelberg.de/~hb3/ellc.html>). Mr. Lemmermeyer informed his teacher professor mr. Peter Roquette with regard to the existence of this proof. It follows copy of letter of Mr. Roquette to Mr. Lemmermeyer in the German where it is proved that the proof of Manin is in actual the same with the proof of Hasse.

29.4.1998

Lieber Herr Lemmermeyer,

besten Dank für die Zusendung Ihres Maninshen Beweises. Bei der Lektüre Ihres Aufschriebs habe ich mich daran erinnern können, daß ich den Maninshen Beweis seinerzeit studiert habe; es ist schon lange her. Und ich kann mich auch an den Eindruck erinnern, den ich damals nach der Lektüre der Arbeit hatte, nämlich daß dies in der Tat im wesentlichen derselbe Beweis wie bei Hasse ist, nur eben unter Benutzung der expliziten Formel für das Additionstheorem der elliptischen Funktionen, was Hasse wegen Charakteristik 2 und 3 vollständig vermeiden wollte (und vermieden hat), und unter Weglassung der strukturellen Deutung der eingeführten Begriffe (was ebenfalls nicht im Sinne von Hasse war).

Allerdings hat natürlich der Maninsche Beweis einen gewissen Wert zum Vortrag in einer Vorlesung für Hörer mit wenigen Vorkenntnissen: das sei ihm gerne zugestanden. (Aufgabe: führe diesen Beweis für Charakteristik 3 und 2 durch!)

Lassen Sie mich vielleicht erklären, wie ich die Sache sehe. Die \mathbb{F}_q -rationalen Punkte von E sind definitionsgemäß gekennzeichnet als die Fixpunkte der Frobenius-Isogenie π von E . Das ist der Grund dafür, daß der Hassende Beweis der Begriff „Isogenie“ benutzt (er sagt: „Meromorphismus“).

Sie $X = (x, y)$ ein allgemeiner Punkt von E (über einem Definitionskörper K , den wir der Einfachheit halber als algebraisch abgeschlossen voraussetzen wollen, was aber nicht notwendig ist). Es ist also $y^2 = x^3 - ax - b$. Es ist $K(X) = K(x, y)$ der Funktionenkörper von E . Jede Isogenie μ wird dann gegeben durch der Punkt $\mu X = (x_\mu, y_\mu)$, der rational ist in $K(X)$. Die „Norm“ von μ wird definiert durch den Körpergrad:

$$N(\mu) = [K(X) : K(\mu X)] \quad (1)$$

In der Regel ist $N(\mu)$ gleich der Anzahl der Punkte im Kern von μ , nämlich dann wenn μ separabel ist (d.h. wenn $K(X)$ separabel ist über $K(\mu X)$). Hierbei muß man aber den unendlich fernen Punkt mitzählen, die Kurve E also projectiv auffassen. Insbesondere folgt

$$N(\pi - 1) = N_q + 1 \quad (2)$$

denn die \mathbb{F}_q -rationalen Punkte bilden den Kern von $\pi - 1$. (Die 1 auf der linken Seite bezeichnet die identische Isogenie; die 1 auf der rechten Seite von (2) ist natürliche Zahl; sie zählt den unendlich fernen Punkt: wie bei Ihnen schreibe ich hier also N_q für die Anzahl der \mathbb{F}_q -rationalen Punkte im endlichen.

Die obige Formel (2) ist die Formel ($\#E(\mathbb{F}_q) = N_q + 1 = d_{-1}$) bei Ihnen. [...]

Der Hassesche Beweis besteht nun darin, die *Normenadditionsformel* zu beweisen:

$$N(\mu + \nu) + N(\mu - \nu) = 2N(\mu) + 2N(\nu) \quad (3)$$

welche zeigt, daß die Norm eine quadratische, *positiv definite* Form definiert auf der additiven Gruppe der Isogenien (wozu auch die uneigentliche Isogenie 0 gezählt wird).

Natürlich genügt es im Hinblick auf (2), diejenige Untergruppe zu betrachten, die aufgespannt wird von der Eins-Isogenie 1 und der Frobenius-Isogenie $\pi = \pi_q$ zu \mathbf{F}_q . Und weiter genügt es, für die Folge $\mu_n = 1 - n\pi$ die Regel

$$N(\mu_{n+1}) + N(\mu_{n-1}) = 2N(\mu_n) + 2 \quad (4)$$

zu zeigen (was ein Spezialfall von (4) ist). Man sieht den Zusammenhang mit der von Ihnen so genannten „Grundrelation“: $d_{n-1} + d_{n+1} = 2d_n + 2$.

Den einzigen neuen Gedanken von Manin sehe ich darin, die Isogenien μ von E darzustellen als $K(x)$ -rationale Punkte der getwisteten Kurve

$$E_\lambda : \lambda z^2 = u^3 + au + b \quad \text{wobei} \quad \lambda = x^3 + ax + b. \quad (5)$$

Zu jeder Isogenie μ von E gehört ein $K(x)$ -rationaler Punkt (u, z) von E_λ , nämlich $u = x_\mu$, $z = y_\mu / y$, und zwar ist dabei $v_\infty(u) < 0$, wobei v_∞ die Bewertung der unendlichen Stelle von $K(x)$ ist, also der negative Grad einer rationalen Funktion. Und umgekehrt: jedem $K(x)$ -rationalen Punkt (u, z) von E_λ entspricht auf diese Weise eine Isogenie μ , derart daß $x_\mu = u$ und $y_\mu = yz$. (Der unendlich ferne Punkt von E_λ gehört zur uneigentlichen Isogenie $\mu = 0$.)

Dabei entspricht der Addition von Isogenien die Addition von Punkten der getwisteten Kurve. Und die Norm einer Isogenie ist

$$N(\mu) = [K(X) : K(\mu X)] = [K(x) : K(x_\mu)] = [K(x) : K(u)] \quad (6)$$

Schreibt man $u = f / g$ mit teilerfremden Polynomen f, g , so ist $v_\infty(u) = -\text{Grad}(f) + \text{Grad}(g) < 0$ und daher

$$[K(x) : K(u)] = \text{Grad}(f) \quad (7)$$

Somit sehen wir, daß die auf Seite 3 Ihres Manuskripts eingeführte Zahl d_n nichts anderes ist als die Norm der zugehörigen Isogenie.

Dieser Zusammenhang erlaubt es Manin, dem Leser den Begriff der Isogenie vorzuenthalten und mit rationalen Punkten der getwisteten Kurve zu rechnen. In Wahrheit ist es aber, wie gesagt, der Hassesche Beweis.

5. Proof of the Basic Identity

For the proof we will need the following lemma:

Lemma 2.5.1 If $(X_n, Y_n) \neq 0$ then $\deg P_n > \deg Q_n$, in particular $X_n \neq 0$.

Proof To prove that the degree of numerator of a rational function $R(x)$ in $\mathbf{F}_p(x)$ is larger than that of its denominator, we formally evaluate $R(x)$ when $x \rightarrow \infty$ and show that $\lim_{x \rightarrow \infty} R(x) = \infty$.

Lemma obviously holds for $n = 0$ because $X_0 = x^p$ and $\deg P_0 = \deg x^p > \deg 1 = \deg Q_0$.

The lemma is also true for those $n > 0$ for which $(X_{n-1}, Y_{n-1}) = 0$ because then $(X_n, Y_n) = (x, 1)$ and $\deg P_n = \deg x > \deg 1 = \deg Q_n$.

Suppose the lemma is true for a particular $n \geq 0$ for which $(X_{n-1}, Y_{n-1}) \neq 0$. We continue by induction.

If we show that the lemma holds for $n + 1$ then we have shown the lemma for any $n \geq 0$. So, we have only to show that if $(X_{n+1}, Y_{n+1}) \neq 0$ then $\deg P_{n+1} > \deg Q_{n+1}$.

It holds:

$$\lambda Y_{n+1}^2 = X_{n+1}^3 + bX_{n+1} + c \quad \eta \quad Y_{n+1}^2 = \frac{X_{n+1}^3 + bX_{n+1} + c}{x^3 + bx + c}.$$

Moreover, because $X_{n+1}(x) = \frac{P_{n+1}(x)}{Q_{n+1}(x)}$ it holds that

$$\lim_{x \rightarrow \infty} X_{n+1}(x) < \infty \Leftrightarrow \deg P_{n+1}(x) \leq \deg Q_{n+1}(x)$$

So, the nominator in the expression of Y_{n+1}^2 above when $x \rightarrow \infty$ is less than ∞ when $\deg P_{n+1}(x) \leq \deg Q_{n+1}(x)$. But then the denominator goes to ∞ .

Hence

$$\lim_{x \rightarrow \infty} Y_{n+1}(x) = 0 \Leftrightarrow \deg P_{n+1}(x) \leq \deg Q_{n+1}(x)$$

so $\lim_{x \rightarrow \infty} Y_{n+1}(x) \neq 0 \Leftrightarrow \deg P_{n+1}(x) > \deg Q_{n+1}(x)$

We will show this with a contradiction.

We assume that $\deg P_{n+1}(x) \leq \deg Q_{n+1}(x)$. Which means that $\lim_{x \rightarrow \infty} Y_{n+1}(x) = 0$.

It holds $(X_{n+1}, Y_{n+1}) = (X_n, Y_n) + (x, 1)$

i.e.

$$(X_{n+1}, -Y_{n+1}) + (X_n, Y_n) + (x, 1) = 0 \quad (2.5.2)$$

This tells us that the points $(X_{n+1}, -Y_{n+1})$, (X_n, Y_n) , $(x, 1)$ lie on a line.

The equation of the line L which intersects (X_n, Y_n) and $(x, 1)$ is

$$L : Y - 1 = \frac{1 - Y_n}{x - X_n} (X - x)$$

Because $(X_{n+1}, -Y_{n+1})$ is also a point on L we get that

$$\begin{aligned} -Y_{n+1} - 1 &= \frac{1 - Y_n}{x - X_n} (X_{n+1} - x) \\ \uparrow \\ Y_{n+1} + 1 &= \frac{1 - Y_n}{x - X_n} (x - X_{n+1}) \\ \uparrow \\ Y_{n+1} &= \frac{1 - Y_n}{x - X_n} (x - X_{n+1}) - 1 \end{aligned}$$

So because $\lim_{x \rightarrow \infty} Y_{n+1}(x) = 0$ we get $\lim_{x \rightarrow \infty} \left[\frac{1 - Y_n}{x - X_n} (x - X_{n+1}) - 1 \right] = 0$ i.e.

$$\lim_{x \rightarrow \infty} \left[\frac{1 - Y_n}{1 - \frac{X_n}{x}} \left(1 - \frac{X_{n+1}}{x} \right) - 1 \right] = 0$$

but because $\deg P_{n+1}(x) \leq \deg Q_{n+1}(x)$ we get $\deg P_{n+1}(x) < \deg Q_{n+1}(x) + 1$ i.e. $\deg P_{n+1}(x) < \deg [xQ_{n+1}(x)]$ so

$$\lim_{x \rightarrow \infty} \frac{X_{n+1}}{x} = \lim_{x \rightarrow \infty} \frac{P_{n+1}}{xQ_{n+1}} = 0 \quad (2.5.3)$$

Thus,

$$\lim_{x \rightarrow \infty} \frac{1 - Y_n}{1 - \frac{X_n}{x}} = 1 \quad (2.5.4)$$

From the addition formula (2.4.6) we get

$$X_{n+1} = \left(\frac{1 - Y_n}{x - X_n} \right)^2 (x^3 + bx + c) - (x + X_n)$$

We get

$$\frac{X_{n+1}}{x} = \left(\frac{1 - Y_n}{1 - \frac{X_n}{x}} \right)^2 \left(1 + \frac{b}{x^2} + \frac{c}{x^3} \right) - 1 - \frac{X_n}{x} \quad (2.5.5)$$

Hence by (2.5.3) and (2.5.5) we get:

$$\lim_{x \rightarrow 0} \left[\left(\frac{1 - Y_n}{1 - \frac{X_n}{x}} \right)^2 \left(1 + \frac{b}{x^2} + \frac{c}{x^3} \right) - 1 - \frac{X_n}{x} \right] \rightarrow 0$$

But by (2.5.4) we get

$$\begin{aligned} & \lim_{x \rightarrow 0} \left[\left(\frac{1 - Y_n}{1 - \frac{X_n}{x}} \right)^2 \left(1 + \frac{b}{x^2} + \frac{c}{x^3} \right) - 1 - \frac{X_n}{x} \right] = \\ & = \lim_{x \rightarrow 0} \left[\left(\frac{1 - Y_n}{1 - \frac{X_n}{x}} \right)^2 \left(1 + \frac{b}{x^2} + \frac{c}{x^3} \right) \right] - 1 - \lim_{x \rightarrow 0} \frac{X_n}{x} = \\ & = 1 - 1 - \lim_{x \rightarrow 0} \frac{X_n}{x} = \\ & = - \lim_{x \rightarrow 0} \frac{X_n}{x} \\ & \text{i.e. } \lim_{x \rightarrow 0} \frac{X_n}{x} \rightarrow 0. \end{aligned}$$

But that means $\deg P_n(x) < \deg [xQ_n(x)]$. So $\deg P_n(x) \leq \deg Q_n(x)$, and we get a contradiction because we assumed that $\deg P_n(x) > \deg Q_n(x)$.

This contradiction proves the lemma for every $n \geq 0$. The induction for $n \leq 0$ is carried out similarly.

We now prove the basic identity: $\mathbf{d}_{n-1} + \mathbf{d}_{n+1} = 2\mathbf{d}_n + 2$

1. If one of (X_{n-1}, Y_{n-1}) , (X_n, Y_n) , (X_{n+1}, Y_{n+1}) is zero the identity is trivial.

In fact, by using (2.5.2) we get:

- i. If $(X_n, Y_n) = 0$ then $(X_0, Y_0) + n(x, 1) = 0$.

Then,

$$(X_{n+1}, Y_{n+1}) = (X_0, Y_0) + (n+1)(x, 1) = (X_0, Y_0) + n(x, 1) + (x, 1) = 0 + (x, 1) = (x, 1).$$

$$\text{Thus, } X_{n+1} = x, Y_{n+1} = 1 \text{ and } (X_{n-1}, Y_{n-1}) = (X_n, Y_n) - (x, 1) = -(x, 1) = (x, -1)$$

$$\text{So } X_{n-1} = x, Y_{n-1} = -1$$

$$\text{Finally, } d_n = 0 \text{ and } d_{n-1} = d_{n+1} = 1$$

- ii. If $(X_{n-1}, Y_{n-1}) = 0$ then $(X_n, Y_n) = (X_0, Y_0) + n(x, 1)$

$$\text{i.e. } (X_n, Y_n) = (X_0, Y_0) + (n-1)(x, 1) + (x, 1) = (X_{n-1}, Y_{n-1}) + (x, 1) = (x, 1)$$

$$\text{So } d_{n-1} = 0, d_n = 1.$$

By the addition formula (2.4.7) we get:

$$\begin{aligned} (X_{n+1}, Y_{n+1}) &= (X_n, Y_n) + (x, 1) = 2(x, 1). \quad X_{n+1} = \frac{(3x^2 + b)^2}{4(x^3 + bx + c)} - 2x = \\ &= \frac{9x^4 + 6bx^2 + b^2 - 8x^4 - 8bx^2 - 8cx}{4(x^3 + bx + c)} = \\ &= \frac{x^4 - 2bx^2 - 8cx + b^2}{4(x^3 + bx + c)}. \end{aligned}$$

We calculate the maximum common divisor of the nominator and the denominator:

- If $b \neq 0$ then

$$\begin{aligned} x^4 - 2bx - 8cx + b^2 &= x(x^3 + bx + c) + (-3bx^2 - 9cx + b^2) \\ x^3 + bx + c &= \left(-\frac{1}{3b}x + \frac{c}{b^2}\right)(-3bx^2 - 9cx + b^2) + \left(\frac{4b}{3} + \frac{9c^2}{b^2}\right)x \\ -3bx^2 - 9cx + b^2 &= (-3bx - 9c) \left(\frac{4b}{3} + \frac{9c^2}{b^2}\right)^{-1} \left(\frac{4b}{3} + \frac{9c^2}{b^2}\right)x + b^2 \end{aligned}$$

b^2 is a unit on the ring $\mathbb{F}_p(x)$ so $x^4 - 2bx - 8cx + b^2$ and $x^3 + bx + c$ are co prime.

- If $b = 0$ then $c \neq 0$ and

$$\begin{aligned} x^4 - 2bx - 8cx + b^2 &= x^4 - 8cx, \\ x^3 + bx + c &= x^3 + c \end{aligned}$$

So:

$$\begin{aligned} x^4 - 8cx &= x(x^3 + c) - 9cx \\ x^3 + c &= -\frac{1}{9c}x^2(-9cx) + c \end{aligned}$$

c is a unit on the ring $\mathbb{F}_p(x)$ so $x^4 - 8cx$ and $x^3 + c$ are co prime.

In both cases $d_{n+1} = 4$.

So $d_{n-1} + d_{n+1} = 0 + 4 = 2 + 2 = d_n + 2$ and the basic identity holds.

iii. If $(X_{n+1}, Y_{n+1}) = 0$ then $(X_n, Y_n) = -(x, 1) = (x, -1)$

So $d_{n+1} = 0, d_n = 1$.

In a similar way as above, by (2.4.7) we get:

$(X_{n-1}, Y_{n-1}) = -2(x, 1) = 2(x, -1)$ and

$$X_{n-1} = \frac{(3x^2 + b)^2}{4(x^3 + bx + c)} - 2x = \frac{x^4 - 2bx^2 - 8cx + b^2}{4(x^3 + bx + c)}.$$

Thus $d_{n-1} = 4$ and $4 + 0 = 2 \cdot 1 + 2$ which holds.

2. Let now $(X_{n-1}, Y_{n-1}) \neq 0, (X_n, Y_n) \neq 0, (X_{n+1}, Y_{n+1}) \neq 0$.

By (2.5.2) we get

$$(X_n, Y_n) = (X_{n-1}, Y_{n-1}) + (x, 1) \text{ i.e.}$$

$$(X_{n-1}, Y_{n-1}) = (X_n, Y_n) - (x, 1) \text{ i.e.}$$

$$(X_{n-1}, Y_{n-1}) = (X_n, Y_n) + (x, -1)$$

By the addition law (2.4.6) and with $X_n = \frac{P_n}{Q_n}$ we get

$$\begin{aligned} X_{n-1} &= \lambda \frac{(Y_n + 1)^2}{(X_n - x)^2} - (x + X_n) = \\ &= \frac{\lambda (Y_n + 1)^2 - (x + X_n)(X_n - x)^2}{(X_n - x)^2} = \\ &= \frac{\lambda (Y_n + 1)^2 - \left(x + \frac{P_n}{Q_n}\right) \left(\frac{P_n}{Q_n} - x\right)^2}{\left(\frac{P_n}{Q_n} - x\right)^2} = \\ &= \frac{Q_n^3 \lambda (Y_n + 1)^2 - (xQ_n + P_n)(P_n - xQ_n)^2}{Q_n (xQ_n - P_n)^2} = \\ &= \frac{\lambda Q_n^2 (Y_n + 1)^2 - (x + X_n)(P_n - xQ_n)^2}{(xQ_n - P_n)^2} = \end{aligned} \tag{2.5.6}$$

$$\begin{aligned}
&= \frac{\lambda Q_n^2 (Y_n^2 + 2Y_n + 1) - (x + X_n)(P_n^2 - xP_nQ_n + x^2Q_n^2)}{(xQ_n - P_n)^2} = \\
&= \frac{Q_n^2 (\lambda Y_n^2 + 2\lambda Y_n + \lambda) - x(P_n^2 - xP_nQ_n + x^2Q_n^2) - X_n(P_n^2 - xP_nQ_n + x^2Q_n^2)}{(xQ_n - P_n)^2} = \\
&= \frac{Q_n^2 (X_n^3 + bX_n + c + 2\lambda Y_n + \lambda) - xP_n^2 + x^2P_nQ_n - x^3Q_n^2 - X_nP_n^2 + xP_n^2 - x^2P_nQ_n}{(xQ_n - P_n)^2} = \\
&= \frac{Q_n^2 X_n^3 + Q_n^2 (bX_n + c + 2\lambda Y_n + \lambda) - xP_n^2 - x^3Q_n^2 - X_nP_n^2 + xP_n^2}{(xQ_n - P_n)^2} = \\
&= \frac{bP_nQ_n + cQ_n^2 + 2\lambda Y_nQ_n^2 + \lambda Q_n^2 - xP_n^2 - x^3Q_n^2 + xP_n^2}{(xQ_n - P_n)^2} = \\
&= \frac{(xQ_n + P_n)(xP_n + bQ_n) + (\lambda - x^3 - bx + c - bx)Q_n^2 + 2\lambda Y_nQ_n^2}{(xQ_n - P_n)^2} = \\
&= \frac{(xQ_n + P_n)(xP_n + bQ_n) + 2cQ_n^2 + 2\lambda Y_nQ_n^2}{(xQ_n - P_n)^2}.
\end{aligned}$$

Thus,
$$X_{n-1} = \frac{P_{n-1}}{Q_{n-1}} = \frac{R}{(xQ_n - P_n)^2} \quad (2.5.7)$$

where $R = (xQ_n + P_n)(xP_n + bQ_n) + 2cQ_n^2 + 2\lambda Y_nQ_n^2$.

Similarly

$$(X_{n+1}, Y_{n+1}) = (X_n, Y_n) + (x, 1)$$

so

$$\begin{aligned}
X_{n+1} &= \lambda \frac{(Y_n - 1)^2}{(X_n - x)^2} - (x + X_n) = \frac{\lambda Q_n^2 (Y_n - 1)^2 - (x + X_n)(P_n - xQ_n)^2}{(xQ_n - P_n)^2} \quad (2.5.8) \\
&= \frac{(xQ_n + P_n)(xP_n + bQ_n) + 2cQ_n^2 - 2\lambda Y_nQ_n^2}{(xQ_n - P_n)^2}.
\end{aligned}$$

Thus,
$$X_{n+1} = \frac{P_{n+1}}{Q_{n+1}} = \frac{S}{(xQ_n - P_n)^2} \quad (2.5.9)$$

where $S = (xQ_n + P_n)(xP_n + bQ_n) + 2cQ_n^2 - 2\lambda Y_nQ_n^2$.

In multiplying the above expressions for X_{n-1} and X_{n+1} we obtain

$$\begin{aligned}
X_{n-1} \cdot X_{n+1} &= \frac{P_{n-1} \cdot P_{n+1}}{Q_{n-1} \cdot Q_{n+1}} = \frac{R \cdot S}{(xQ_n - P_n)^4} = \\
&= \frac{[(xQ_n + P_n)(xP_n + bQ_n) + 2cQ_n^2]^2 - 4\lambda^2 \cdot Y_n^2 \cdot Q_n^4}{(xQ_n - P_n)^4} = \\
&= \frac{[(xQ_n + P_n)(xP_n + bQ_n) + 2cQ_n^2]^2 - 4\lambda(X_n^3 + bX_n + c) \cdot Q_n^4}{(xQ_n - P_n)^4} = \\
&= \frac{[(xQ_n + P_n)(xP_n + bQ_n) + 2cQ_n^2]^2 - 4(x^3 + bx + c)(P_n^3Q_n + bP_n^4Q_n^3 + cQ_n^4)}{(xQ_n - P_n)^4} = \\
&= \frac{x^4Q_n^2P_n^2 + 2x^3Q_n^3P_nb + x^2Q_n^4b^2 - 2x^3Q_nP_n^3 + 4x^2Q_n^2P_n^2b + 2xQ_n^3P_nb^2 + P_n^4x^2}{(xQ_n - P_n)^4} + \\
&+ \frac{-2P_n^3xbQ_n + P_n^2b^2Q_n^2 + 4cQ_n^3x^2P_n + 4cQ_n^3b + 4cQ_n^2xP_n^2}{(xQ_n - P_n)^4} + \\
&+ \frac{-4x^3bP_n^4Q_n^3 - 4x^3cQ_n^4 - 4b^2xP_n^4Q_n^3 - 4cP_n^3Q_n - 4cbP_n^4Q_n^3}{(xQ_n - P_n)^4} = \\
&= \frac{(xQ_n - P_n)^2 [(xP_n - bQ_n)^2 - 4cQ_n(xQ_n + P_n)]}{(xQ_n - P_n)^4} = \\
&= \frac{(xP_n - bQ_n)^2 - 4cQ_n(xQ_n + P_n)}{(xQ_n - P_n)^2}. \tag{2.5.10}
\end{aligned}$$

If we show that $Q_{n-1} \cdot Q_{n+1} = k \cdot (xQ_n - P_n)^2$, where $k \in \mathbf{F}_p$ then

$$P_{n-1} \cdot P_{n+1} = k \cdot [(xP_n - bQ_n)^2 - 4cQ_n(xQ_n + P_n)]$$

and then

$$d_{n-1} + d_{n+1} = \deg(P_{n-1} \cdot P_{n+1}) = \deg(x^2 \cdot P_n^2 \cdot k) = \deg(x^2) + \deg(P_n^2) = 2d_n + 2$$

and we have completed the proof of the basic identity.

From the equalities of (2.5.10) we get

$$(xQ_n - P_n)^2 R \cdot S = (xQ_n - P_n)^4 \cdot [(xP_n - bQ_n)^2 - 4cQ_n(xQ_n + P_n)] \acute{u}$$

$$R \cdot S = (xQ_n - P_n)^2 \cdot [(xP_n - bQ_n)^2 - 4cQ_n(xQ_n + P_n)]$$

Thus, $(xQ_n - P_n)^2 \mid R \cdot S$

Hence there exist polynomials $R_1, S_1 \in \mathbf{F}_p[x]$ such that $(xQ_n - P_n)^2 = R_1 \cdot S_1$ where $R_1 \mid R$ and $S_1 \mid S$.

So $X_{n-1} = \frac{P_{n-1}}{Q_{n-1}} \stackrel{(2.5.7)}{=} \frac{R}{(xQ_n - P_n)^2} = \frac{R}{R_1 \cdot S_1} = \frac{\frac{R}{R_1}}{S_1}$ because $\frac{P_{n-1}}{Q_{n-1}}$ is in the lowest form, i.e. that Q_{n-1}, P_{n-1} are co prime, we get that $Q_{n-1} \mid S_1$.

Similarly, $X_{n+1} = \frac{P_{n+1}}{Q_{n+1}} \stackrel{(2.5.9)}{=} \frac{S}{(xQ_n - P_n)^2} = \frac{S}{R_1 \cdot S_1} = \frac{\frac{S}{S_1}}{R_1}$ and Q_{n+1}, P_{n+1} are co prime. So $Q_{n+1} \mid R_1$.

Hence, $Q_{n-1} \cdot Q_{n+1}$ divides $S_1 \cdot R_1 = (xQ_n - P_n)^2$.

So

To show $Q_{n-1} \cdot Q_{n+1} = k \cdot (xQ_n - P_n)^2$ we only have to show that the polynomial $(xQ_n - P_n)^2$ divides $Q_{n-1} \cdot Q_{n+1}$ (1)

Definition 2.5.11 We consider a polynomial $A \neq 0$ and a irreducible polynomial f on the Euclidean field $\mathbf{F}_p[x]$. $v_f(A)$ is defined to be the power to which f appears in the factorization of A into powers of prime (irreducible) polynomials. If $A = 0$ we define $v_f(A) = \infty$. It is clear that $v_f(A) \in \{0, 1, 2, \dots\} \cup \{\infty\}$.

v_f converges on the field of the rational functions $\mathbf{F}_p(x)$:

If $T \in \mathbf{F}_p(x)$ and $T = \frac{A}{B}$ with $A, B \in \mathbf{F}_p[x]$ then the map $v_f(T)$ is defined to be $v_f(T) = v_f(A) - v_f(B)$. Hence, $v_f(T) \in \mathbf{Z} \cup \{\infty\}$.

If $A, B \in \mathbf{F}_p(x)$ and $A \mid B$ then obviously $v_f(A) \leq v_f(B)$ for any irreducible $f \in \mathbf{F}_p[x]$.

So, $A \nmid B$ means that there is an irreducible $f \in \mathbf{F}_p[x]$ such that

$$v_f(A) > v_f(B)$$

Remark 2.5.12 For any $A \in \mathbf{F}_p(x)$ it holds $v_f(A^2) = 2 v_f(A)$

We go back to the proof of the basic identity. Let $(xQ_n - P_n)^2 \nmid Q_{n-1} \cdot Q_{n+1}$ that there is an irreducible $f \in \mathbf{F}_p[x]$ such that

$$v_f((xQ_n - P_n)^2) > v_f(Q_{n-1} \cdot Q_{n+1}) \quad \text{i.e.}$$

$$2 v_f(xQ_n - P_n) > v_f(Q_{n-1} \cdot Q_{n+1}) \quad (2.5.13)$$

To help our calculations we define

$$T := (xP_n - bQ_n)^2 - 4cQ_n(xQ_n + P_n) \quad (2.5.14)$$

So from (2.5.10) we get

$$\frac{P_{n-1} \cdot P_{n+1}}{Q_{n-1} \cdot Q_{n+1}} = \frac{T}{(xQ_n - P_n)^2},$$

i.e.

$$T = \frac{P_{n-1} \cdot P_{n+1} \cdot (xQ_n - P_n)^2}{Q_{n-1} \cdot Q_{n+1}} \quad (2.5.15)$$

From (2.5.15) and (2.5.13) we get that f divides T .

The polynomial f divides $(xQ_n - P_n)^2$ so if we prove that f divide both R and S then, from (2.5.6) and (2.5.8), we get

$$f \mid \lambda Q_n^2 (Y_n + 1)^2 \text{ and } f \mid \lambda Q_n^2 (Y_n - 1)^2$$

If we assume that f divides Q_n , then, because it also divides $xQ_n - P_n$, it divides P_n which is a contradiction because P_n, Q_n are co prime. So, f does not divide Q_n . We get

$$f \mid \lambda (Y_n + 1) \text{ and } f \mid \lambda (Y_n - 1)$$

We now assume that f does not divide λ . Then $f \mid (Y_n + 1)$ and $f \mid (Y_n - 1)$. So f divides $(Y_n + 1) - (Y_n - 1) = 2$, which is a contradiction because f is an irreducible polynomial and not a constant of $\mathbb{F}_p[x]$.

Hence, f divides $\lambda = x^3 + bx + c$.

We write T as a polynomial of P_n :

$$T = (xP_n - bQ_n)^2 - 4cQ_n(xQ_n + P_n) = x^2P_n^2 - 2bxP_nQ_n + b^2Q_n^2 - 4cxQ_n^2 - 4cQ_nP_n = x^2P_n^2 + (-2bx - 4c)Q_nP_n + (b^2 - 4cxQ_n^2)$$
 and we then divide T with $xQ_n - P_n$:

$$T = -(xQ_n - P_n) [x^2P_n + (x^3 - 2bx - 4c)Q_n] + (x^4 - 2bx - 8cx + b^2)Q_n^2.$$

The polynomial f divides both T and the polynomial $xQ_n - P_n$ so it divides the polynomial $(x^4 - 2bx - 8cx + b^2)Q_n^2$. Also, f does not divide Q_n . This shows that f divides $x^4 - 2bx - 8cx + b^2$. But then f divides

$$(3x^3 - 5bx - 27c)(x^3 + bx + c) - (3x^2 + b)(x^4 - 2bx - 8cx + b^2) = -4b^3 - 27c^2 = \Delta,$$

which is the discriminant of the elliptic curve. This is a contradiction because Δ is a non-zero constant.

Thus, we only have to prove that, f divides both R and S to complete the proof.

We have that f divides T and also that T divides RS (see (2.5.10)). So, $f \mid RS$. Hence, $f \mid R$ or $f \mid S$.

Now suppose f divides R but not S (Then proof in the case that $f \mid S$ and $f \nmid R$ is exactly the same).

Because $f \nmid S$ and $f \mid (xQ_n - P_n)^2$, see (2.5.13), we get $v_f(Q_{n+1}) = v_f(Q_{n+1}S) \stackrel{(2.5.9)}{=} v_f((xQ_n - P_n)^2 P_{n+1}) = v_f((xQ_n - P_n)^2) + v_f(P_{n+1}) > 0$. Hence, $f \mid Q_{n+1}$. But then because P_{n+1}, Q_{n+1} are co prime we get $f \nmid P_{n+1}$ i.e.

$$v_f(P_{n+1}) = 0. \quad (2.5.16)$$

Also

$$v_f(Q_{n+1}) = 2 v_f(xQ_n - P_n) \quad (2.5.17)$$

Now we calculate $v_f(T)$.

Because f divides R it also divides $T = R \cdot S \cdot (xQ_n - P_n)^2$. So $v_f(T) > 0$. Also, due to (2.5.15):

$$0 < v_f(T) = v_f\left(\frac{P_{n-1} \cdot P_{n+1} \cdot (xQ_n - P_n)^2}{Q_{n-1} \cdot Q_{n+1}}\right) = v_f(P_{n-1} \cdot P_{n+1} \cdot (xQ_n - P_n)^2) - v_f(Q_{n-1} Q_{n+1}) = v_f(P_{n-1}) + v_f(P_{n+1}) + 2 v_f(xQ_n - P_n) - v_f(Q_{n-1}) - v_f(Q_{n+1}).$$

From (2.5.16) and (2.5.17) we get

$$0 < v_f(T) = v_f(P_{n-1}) - v_f(Q_{n-1})$$

Hence,

$$v_f(P_{n-1}) > v_f(Q_{n-1})$$

This means that $v_f(P_{n-1}) > 0$. So f divides P_{n-1} . Because P_{n-1} and Q_{n-1} are co prime, we get that f does not divide Q_{n-1} , i.e.

$$v_f(Q_{n-1}) = 0. \quad (2.5.18)$$

From (2.5.17) and (2.5.18):

$$v_f(Q_{n-1} Q_{n+1}) = v_f(Q_{n-1}) + v_f(Q_{n+1}) = 0 + 2v_f(xQ_n - P_n) = 2v_f(xQ_n - P_n).$$

This is a contradiction because we assumed, see (2.5.13),

$$2 v_f(xQ_n - P_n) > v_f(Q_{n-1} \cdot Q_{n+1}).$$

Hence, we proved that $(xQ_n - P_n)^2 \mid (Q_{n-1} \cdot Q_{n+1})$. This completes the proof of the basic identity.

Chapter III

Algebraic Curves and Coding Theory

1. Elements of coding theory

Definition 3.1.1

- (i) We call **alphabet** the finite set of symbols (which we usually call **letters**) we use to form a message. Our alphabet, in this chapter, will be the finite field \mathbf{F}_q .
- (ii) A **k-message** consists of a sequence of letters from our alphabet of length k i.e. it is in the form a_1, a_2, \dots, a_k with $a_i \in \mathbf{F}_q$.
- (iii) The corresponding **codeword** x of a k -message is a sequence of length n . So, it written in the form $x = x_1, x_2, \dots, x_n$ with $x_i \in \mathbf{F}_q$ and $n \geq k$. Hence we (mostly) assume that $x_1 = a_1, x_2 = a_2, \dots, x_k = a_k$ and the remaining $n - k$ elements $(x_{k+1}, x_{k+2}, \dots, x_n)$ are called **check symbols** or **control symbols**.

Notation 3.1.2 Codewords will be written in one of the forms x or x_1, x_2, \dots, x_n or (x_1, x_2, \dots, x_n) or $x_1x_2\dots x_n$.

Definition 3.1.3 We shall call **received vector** (or **received message**) the vector $y = y_1, y_2, \dots, y_n$ we receive. The vector y is, in general, different from the vector x which was sent. The vector $e := y - x = e_1e_2\dots e_n$ is called **error vector** (or just **error**).

Definition 3.1.4 An **n-code** C is a subset of \mathbf{F}_q^n . More precisely the code will be called **(n, k)-code**, where k is the length of the message we encode. If the code C is an \mathbf{F}_q -vector subspace of \mathbf{F}_q^n we will call it **(n, k)-linear code**. The elements of C are the codewords.

Examples 3.1.5

- (i) Let $C = \{000, 001, 010, 011\}$ is a subset of \mathbf{F}_2^3 . C consists of exactly all the codewords which have 0 as their first coordinate and easily we can see that C is a linear 3-code.
- (ii) Also, $C = \{00, 11, 22\}$ is a linear 2-code of \mathbf{F}_3^2 .

After receiving y , we have to determine which codeword was sent. We choose a codeword from the set C that is closest to y that means it has the minimum Hamming distance from y . This decoding method is called the “**nearest neighbor decoding**”.

Definition 3.1.6 The **Hamming distance** $d(x, y)$ between two vectors x, y in \mathbf{F}_q^n , with

$$x = x_1, x_2, \dots, x_n \text{ and } y = y_1, y_2, \dots, y_n,$$

is the number of coordinates in which x and y differ i.e.

$$d(x, y) = \#\{i \in \mathbf{N}, 1 \leq i \leq n \mid x_i \neq y_i\}$$

Definition 3.1.7 The **Hamming weight** $w(x)$ of a vector $x = x_1, x_2, \dots, x_n$ in \mathbf{F}_q^n is the number of nonzero coordinates of x i.e.

$$w(x) = \#\{i \in \mathbf{N}, 1 \leq i \leq n \mid x_i \neq 0\}$$

Obviously, $w(x) = d(x, \mathbf{0})$, where $\mathbf{0} = 00\dots 0$.

Example 3.1.8 Let $C \subseteq \mathbf{F}_3^4$.

The Hamming weight of 1201 is $w(1201) = 3$.

The Hamming distance of 1201 and 2211 is $d(1201, 2211) = 2$.

Remark 3.1.9 The Hamming distance $d(C)$ is a metric on \mathbf{F}_q^n and the Hamming weight w is a norm on \mathbf{F}_2^n .

Definition 3.1.10 If $C \subseteq \mathbf{F}_q^n$ is an (n, k) -code, **the minimum distance $d_{\min}(C)$ of the code** is

$$d_{\min}(C) = \min_{\substack{u, v \in C \\ u \neq v}} d(u, v)$$

So, after receiving y , we have to search through all q^k codewords and find one which is closest to y . Obviously this procedure is impossible for large k and one of the aims of coding theory is to find codes with faster decoding algorithms.

The next result will show that for every linear code we can calculate its minimum distance from the Hamming weight of the codewords.

One of the most important properties of the linear codes is the following

Proposition 3.1.11 Let C be a linear (n, k) -code. The minimum distance of C is equal to the least weight of all nonzero codewords.

Proof Let w be the minimum Hamming weight of a non-zero codeword. Let $x \in C$ be a codeword of Hamming weight w . It holds that $d(x, \mathbf{0}) = w(x) = w$. So, $w \geq d_{\min}(C)$. Let now u and v be a pair of codewords of C with distance such that $d(u, v) = d_{\min}(C)$. C is a linear code so $u - v$ is also a codeword. $u - v$ has weight equal to $d_{\min}(C)$. Then, $d_{\min}(C) \geq w$. Thus, $d_{\min}(C) = w$.

Definition 3.1.12 The set $S_r(x) := \{y \in \mathbf{F}_q^n \mid d(x, y) \leq r\}$ is called **the sphere of radius r about $x \in \mathbf{F}_q^n$** .

Example 3.1.13 Let $C = \mathbf{F}_2^3$ then the sphere of radius 1 about 100 is

$$S_1(100) = \{100, 000, 110, 101\}.$$

Aim By taking spheres of some radius r about a codeword we want, if possible, their union to cover the space \mathbf{F}_q^n in order to decode all messages we get and on the other hand the radius r must be small enough so that two spheres do not intersect not even in a single point in order to decode in a single point.

So r must be smaller than $\frac{1}{2} d_{\min}(C)$.

The importance of the minimum distance is given by the

Proposition 3.1.14 A linear code C with minimum distance $d_{\min}(C) = d$. C can **detect** up to $d - 1$ errors and **correct** e errors for every e such that $2e + 1 \leq d$.

Proof We assume that we received a message y with distance f from the codeword x , where $f \leq d - 1$. This means we have f errors during transmission. Because the minimum distance of C is d we note that y can't be a codeword. This means that the code C can detect up to $d - 1$ errors.

If the message y has distance e from the codeword x and $2e + 1 \leq d$ then there isn't any other codeword closer to y because if $d(y, x_1) \leq e$ for some codeword x_1 then we would get that

$$d(x, x_1) \leq d(x, y) + d(y, x_1) \leq e + e < d$$

which is a **contradiction**, because the minimum distance of the code C is d . So, there is a unique codeword that is closer to y and the code can correct e errors in this case.

One of the main problems of coding theory is not merely to minimize errors, but to do so without reducing the **information rate** $\frac{k}{n}$.

The main problem of the Coding Theory is the following:

Given d, n natural numbers find the maximum number of vectors, let $A_q(n, d)$, in the vector space \mathbf{F}_q^n which are at distances more or equal d . Of course, if it is possible, find these vectors.

For $q = 2$ we denote $A(n, d) = A_2(n, d)$.

The following table gives some values of $A_2(n, d)$ for $d = 3$

N	3	4	5	6	7	8	9	10
$A_2(n, 3)$	2	2	4	8	16	20	40	unknown, between 72 and 79

This problem is known as the “discrete sphere packing problem” (Conway and Sloane, 1988)

Definition 3.1.15 A code that corrects t errors is called **t-error-correcting**, and a code that detects e errors is called **e-error-detecting**.

Let now C be a code over \mathbf{F}_q of length n with M codewords. Suppose C is t -error correcting. There are $(q-1)^m \binom{n}{m}$ vectors of \mathbf{F}_q^n of length n and weight m over \mathbf{F}_q . If $c \in C$ then within or on the sphere $S_t(c)$ there are totally $1 + (q-1) \binom{n}{1} + \dots + (q-1)^t \binom{n}{t}$ vectors of \mathbf{F}_q^n .

Theorem 3.1.16 (Hamming Bound) The parameters q, n, t, M of a t -error-correcting code C defined over \mathbf{F}_q of length n with M codewords satisfy the inequality

$$M \left(1 + (q-1) \binom{n}{1} + \dots + (q-1)^t \binom{n}{t} \right) \leq q^n.$$

If all the vectors of \mathbf{F}_q^n are within or on spheres of radius t around codewords of a linear (n, k) code then we obtain a special class of codes:

Definition 3.1.17 A t -error-correcting code over \mathbf{F}_q is called **perfect** if equality holds in theorem 3.1.16.

If C is a code as in theorem 3.1.16 with $d_{\min}(C) = d = 2t + 1$, then deleting the last $d - 1$ symbols still gives a code in which all words are different. Since this code has length $n - d + 1$, we get

Theorem 3.1.18 (Singleton Bound) If a code $C \subseteq \mathbf{F}_q^n$ has minimum distance d , then $|C| \leq q^{n-d+1}$ i.e. $k \leq n - d + 1$.

Definition 3.1.19 A code is called **maximum distance separable** or briefly **MDS code** if equality holds in theorem 3.1.18.

Example 3.1.20 We consider the code

$$C = \{000000, 001011, 010101, 011110, 100110, 101101, 110011, 111000\} \subseteq \mathbf{F}_2^6$$

where $d_{\min}(C) = 3$. We have $M = 8, q = 2, n = 6, d = 3, t = 1$. The Hamming bound gives the inequality $8 \left(1 + \binom{6}{1} \right) \leq 2^6$, i.e. < 64 . This means that only $64 - 56 = 8$ 6-

tuples of \mathbf{F}_2^6 lie outside of spheres and cannot be corrected properly (this is also clear from the fact that we have eight nonintersecting spheres with seven elements each). One example of these eight incoming words that cannot be corrected properly is 100100 which has a distance more or equal to 2 from all the codewords. The Singleton bound yields $8 \leq 2^4 = 16$, so this code is not MDS.

Suppose now that the check symbols can be obtained from the k -message symbols in such a way that the codewords x satisfy the system of linear equations

$$Hx^T = 0,$$

where H is a given $(n - k) \times n$ matrix with elements in the field \mathbf{F}_q . The **standard form** for H is $[A \mid I_{n-k}]$ where A is an $(n - k) \times k$ matrix and I_{n-k} the $(n - k) \times (n - k)$ identity matrix.

In general, we define

Definition 3.1.21 Let H be an $(n - k) \times n$ matrix of rank $n - k$ with elements in \mathbf{F}_q . The set of all n -dimensional vectors x satisfying the equation $Hx^T = 0$ over \mathbf{F}_q is called the **linear (block) matrix C** over \mathbf{F}_q of **(block) length n** . The matrix H is called the **parity-check matrix** of the code. C is also called a linear (n, k) -code. If H is of the form $[A \mid I_{n-k}]$ then the first k symbols in the codeword x are the message symbols, and the last $n - k$ symbols on x are the check symbols. C is then also called **systematic linear (n, k) -code** and H is said to be in **standard form**. If $q = 2$, then C is a **binary code**.

Remark 3.1.22 The set C of solutions x of $Hx^T = 0$ (i.e. the null space of H) forms a subspace of \mathbf{F}_q^n of dimension k . Since the codewords form an additive group, C is also called a **group code**.

Example 3.1.23 (Repetition Code) If each codeword of a code consists of only one message symbol $a_1 \in \mathbf{F}_q$ and the $n - 1$ check symbols $x_2 = \dots = x_n$ are all equal to a_1 (a_1 is repeated $n - 1$ times) then we obtain a binary $(n, 1)$ -code with parity-check matrix

$$H = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix}$$

There are only two codewords in this code, namely $00\dots 0$ and $11\dots 1$.

In repetition codes we can, of course, also consider codewords with more than one message symbol. If we transmit a message of length k three times and we compare corresponding “coordinates” x_i, x_{k+i}, x_{2k+i} of the codeword

$$x_1 \dots x_i \dots x_k x_{k+1} \dots x_{k+i} \dots x_{2k} x_{2k+1} \dots x_{2k+i} \dots x_{3k},$$

then a “majority decision” decides which k -message has been sent, for example if $x_i = x_{k+i} \neq x_{2k+i}$, then x_i is more likely from x_{2k+i} to have been transmitted. It is, however, often impractically, impossible, or too expensive to send the original message more than once.

We have seen that in a systematic code, a message $\mathbf{a} = a_1, \dots, a_k$ is encoded as codeword $\mathbf{x} = x_1, \dots, x_n$ with $x_1 = a_1, x_2 = a_2, \dots, x_k = a_k$. The check equations $[\mathbf{A} \mid \mathbf{I}_{n-k}]\mathbf{x}^T = 0$ are given by

$$\begin{pmatrix} x_{k+1} \\ \vdots \\ x_n \end{pmatrix} = -\mathbf{A} \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} = -\mathbf{A} \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix}$$

thus we obtain

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{bmatrix} \mathbf{I}_k \\ -\mathbf{A} \end{bmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix}.$$

We transpose and write this equation as

$$(x_1, \dots, x_n) = (a_1, \dots, a_k) [\mathbf{I}_k \mid -\mathbf{A}^T].$$

Definition 3.1.24 The matrix $\mathbf{G} = [\mathbf{I}_k \mid -\mathbf{A}^T]$ is called **(canonical) generator matrix** (or **canonical basic matrix** or **encoding matrix**) of a linear (n, k) -code with parity-check matrix $\mathbf{H} = [\mathbf{A} \mid \mathbf{I}_{n-k}]$ in standard form.

It holds $\mathbf{GH}^T = 0$.

Examples 3.1.25

(1) Let $\mathbf{G} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$.

The linear $(2,3)$ -code $\mathbf{C} \subseteq \mathbf{F}_2^3$ consists of all the combinations of the two lines:

$$000, 101, 011, 110$$

The codewords can be described as vectors of the form \mathbf{uG} , where $\mathbf{u} = 00, 01, 10, 11$. Each codeword, other than the zero codeword, has weight equal to 2. This means that the code detects up to 1 error and does not correct any of them.

(2) Let $\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$.

From the three lines of the matrix we obtain a $(3, 6)$ -code $\mathbf{C} \subseteq \mathbf{F}_2^6$ which consists of 8 codewords:

$$000000, 100110, 010101, 001011, 110011, 011110, 101101, 111000$$

Like before each codeword x can be described as a vector of the form uG , where $u = u_1u_2u_3$ with u_i in \mathbf{F}_2 .

There are four codewords of weight 3, three codewords of weight 4 and one codeword of weight 0. The minimum distance of the code is 3, so it detects 2 errors and can detect one.

$$(3) \text{ Let } G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 2 & 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 2 & 2 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

The code $C \subseteq \mathbf{F}_3^{12}$ consists of codewords where each codeword x can be described as a vector of the form $x = uG$, where $u = u_1u_2u_3u_4u_5u_6$ with u_i in \mathbf{F}_3 .

The minimum distance of the code is less or equal to 5, because there is a line in the matrix G of weight 5. It can be proven that the code has minimum distance exactly 5. This code is called **Golay code**.

For more information about Golay codes the interested reader can see [MS], chapter20.

Theorem 3.1.26 Let G be a generator matrix of a linear code C . Then the rows of G form a basis of C .

Proof The k rows of the matrix G are linearly independent due to the definition of the generator matrix of a linear code. If r is a row-vector of G it holds that $rH^T = 0$ i.e. $Hr^T = 0$ for all $r \in C$. $\dim C$ is the dimension of the null space of H and it is equal to $n - \text{rank}(H) = k$. So, the k rows of G form a basis of C .

A code can have several parity-check matrices and generator matrices. Any $k \times n$ matrix whose row space is equal to C can be taken as a generator matrix of C .

If the “generator matrix” H is not on the standard form we can perform row operations in H to transform it in a matrix in the form $[I_k | -A^T]$ without changing the null space of H , i.e. the code C . Then we permute the coordinates to get H finally into standard form H' . The corresponding code C' is equivalent to C in the following sense:

Definition 3.1.27 Two codes C and C' of the same length n are called **equivalent** if there is a permutation π of the set $\{1, 2, \dots, n\}$ such that

$$(x_1, \dots, x_n) \in C \Leftrightarrow (x_{\pi(1)}, \dots, x_{\pi(n)}) \in C'$$

We form the generator matrix G' of C' and then perform the inverse permutation π^{-1} of the coordinates.

We now state a definition that is useful in the next paragraph:

Definition 3.1.28 A linear code of length n , dimension k , and minimum distance d is called an **(n, k, d) -code**.

Let, $u = u_1, \dots, u_n$ and $v = v_1, \dots, v_n$ be two vectors in the vector space \mathbf{F}_q^n and let $u \cdot v = u_1v_1 + \dots + u_nv_n$ denote the dot product of u and v over \mathbf{F}_q^n . If $u \cdot v = 0$ then u and v are called **orthogonal**.

Definition 3.1.29 Let C be a linear (n, k) -code over the field \mathbf{F}_q . The **dual (or orthogonal) code** C^\perp of C is defined by

$$C^\perp = \{u \mid uv = 0 \text{ for all } v \in C\}$$

Since C is a k -dimensional subspace of the n -dimensional vector space \mathbf{F}_q^n the orthogonal complement is of dimension $n - k$ and an $(n, n - k)$ code. It can be shown that if the code C has a generator matrix G and parity-check matrix H , then C^\perp has generator matrix H and parity-check matrix G . Orthogonality of the two codes can be expressed by the relation $GH^T = HG^T = 0$. We now summarize some of the simple properties of linear codes.

Remark 3.1.30 Let $\text{mld}(H)$ be the minimal number of linearly dependent columns of H . Since any $\text{rank}(H) + 1$ rows of H are linearly dependent it holds that $\text{mld}(H) \leq \text{rank}(H) + 1$ for all matrices H .

Theorem 3.1.31 Let H be a parity-check matrix of an (n, k, d) -code C with $n > k$. Then the following hold:

- (i) $\dim C = k = n - \text{rank}(H)$
- (ii) $d = \text{mld}(H)$
- (iii) $d \leq n - k + 1$.

Proof (i) is clear and (iii) occurs from (ii) and the previous remark. In order to see (ii) suppose H has the columns s_1, \dots, s_n . If $c = (c_1, \dots, c_n) \in C$ of weight w then because

$$Hc^T = c_1s_1 + \dots + c_ns_n$$

It holds that $c_1s_1 + \dots + c_ns_n = 0$. We also have that c is non-zero in w coordinates so some w and not less rows of H are linearly dependent. Thus, $\text{mld}(H) = w$. Applying proposition 3.1.11 we have that the minimum distance d of the code is equal with the weight of c and we proved (ii).

To verify the existence of linear (n, k) -codes with minimum distance d over \mathbf{F}_q it suffice to show that there exists an $(n - k) \times n$ matrix H with $\text{mld}(H) = d$, for given n, k, d, q .

Theorem 3.1.32 (Gilbert – Varshamov Bound)

If

$$q^{n-k} > \sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i$$

then we can construct a linear (n, k) -code over \mathbf{F}_q with minimum distance more or equal than d . (See [LP], chapter 4, theorem 17.14, page 196).

Proof We construct an $(n - k) \times n$ matrix H for such a code. Let the first column of H be any nonzero $(n - k)$ -tuple over \mathbf{F}_q . The second column is any $(n - k)$ -tuple over \mathbf{F}_q which is not a scalar multiple of the first column. Suppose $j - 1$ columns have been chosen so that any $d - 1$ of them are linearly independent. There are at most $\sum_{i=0}^{d-2} \binom{j-1}{i} (q-1)^i$ vectors obtained by taking linear combinations of $d - 2$ or fewer of these $j - 1$ columns. If the inequality of the theorem holds, it will be possible to choose a j -th column which is linearly independent of any $d - 2$ of the first $j - 1$ columns. This construction can be carried out in such a way that H has rank $n - k$. Because no $d - 1$ columns of H are linearly dependent the resulting code has minimum distance more or equal to d .

Without proof we state the following

Theorem 3.1.33 (Plotkin Bound) If there is a linear code of length n with M codewords and minimum distance d over \mathbf{F}_q , then

$$d \leq n \frac{M(q-1)}{(M-1)q}$$

(See [LP], chapter 4, theorem 17.15, page 197).

In order to achieve better coding we can construct concatenated codes concatenating two codes as following:

Let C_1 be an (n_1, k_1, d_1) -code and C_2 an (n_2, k_2, d_2) -code. We assume that the message we want to encode is the $\alpha = \alpha_1, \dots, \alpha_{k_1}$ where $\alpha_i = \beta_{i_1}, \beta_{i_2}, \dots, \beta_{i_{k_2}} \mu \epsilon \alpha_i \in \text{GF}(2^{k_2})$. With the use of the code C_1 we encode α to $c = c_1, \dots, c_{n_1}$ where $c_i \in \text{GF}(2^{k_2})$ is of the form $c_i = \gamma_{i_1}, \gamma_{i_2}, \dots, \gamma_{i_{k_2}}$. Then we encode each c_i with the use of the code C_2 to $y_i = y_{i_1}, y_{i_2}, \dots, y_{i_{n_2}}$. Thus, we have encoded the message α we want

to send with the use C , which is the concatenation of code C_1 and C_2 , to the codeword $c = (y_{1_1}, y_{1_2}, \dots, y_{1_{n_1}})(y_{2_1}, y_{2_2}, \dots, y_{2_{n_2}}) \dots (y_{n_1}, y_{n_2}, \dots, y_{n_{n_2}})$.

The following holds

Proposition 3.1.34 The minimum distance of the code C that we described above is at least $d_1 \cdot d_2$.

2. Codes defined using algebraic curves

In the following \mathbf{F} is the algebraic closure of the field \mathbf{F}_q . If X is an affine curve defined over \mathbf{F}_q , we denote with $\mathbf{F}_q[X]$ the coordinate ring of X and with $\mathbf{F}_q(X)$ the function field of X , which is also the quotient field of $\mathbf{F}_q[X]$. We assume that the curve is absolutely irreducible, this means it remains irreducible as a curve over \mathbf{F} . We make similar assumptions for the projective curves. Note also that for all polynomials $F \in \mathbf{F}_q[X, Y]$ it holds $F(x_1, y_1)^q = F(x_1^q, y_1^q)$. This means that if (x_1, y_1) is a zero of F , over the field \mathbf{F}_q , then (x_1^q, y_1^q) is also a zero of F .

If the curve X is defined over \mathbf{F}_q and P is a point on the curve then due to the last remark the point $\text{Fr}(P)$ is also on the curve X (see proposition 2.3.1).

Definition 3.2.1 A **divisor** D on a curve X is a formal sum of the form $D = \sum_{P \in X} n_P P$

where $n_P \in \mathbf{Z}$ and $n_P = 0$ for all but a finite number of points P on X .

We define the sum on divisors, naturally, in coordinates.

Definition 3.2.2 The **support** of a divisor is the set of points with nonzero coefficient n_P . A divisor is called **effective** if all coefficients n_P are nonnegative.

Definition 3.2.3 The **degree** $\deg(D)$ of a divisor $D = \sum_{P \in X} n_P P$ is called the, finite, sum

$$\deg(D) = \sum_{P \in X} n_P .$$

Definition 3.2.4 A divisor D on a curve X is called **rational** if the coefficients of P and $\text{Fr}(P)$ are the same for all P on X .

Definition 3.2.5 Let D be a divisor on a curve X . We define the \mathbf{F} -vector space of finite dimension

$$L(D) := \{ f \in \mathbf{F}(X)^* \mid \langle f \rangle + D \geq 0 \} \cup \{0\}.$$

With $\langle f \rangle$ we denote the principal divisor of a rational function f , which is defined as $\langle f \rangle = \sum_{P \in X} n_P P$ and the coefficient n_P is non-zero to the points P in which f has a **zero** with multiplicity n_P , $n_P > 0$ and to the points which f has **pole** of order $-n_P$, $n_P < 0$ and in all other points the coefficient n_P is equal to zero.

We state without proof the

Theorem 3.2.6 The degree of a principal divisor $\langle f \rangle$ of a rational function f is always 0. (For the proof see [Ho], theorem 2.32, page 885)

We can modify the definition 3.2.5 to define the \mathbf{F} -vector space $L(D)_{\text{rat}}$. Of the rational divisors on a curve X :

Definition 3.2.7 Let D be a **rational** divisor on a curve X . We define the \mathbf{F} -vector space of finite dimension

$$L(D)_{\text{rat.}} := \{f \in \mathbf{F}_q(X)^* \mid \langle f \rangle + D \geq 0\} \cup \{0\}.$$

With these change all the stated theorems for $L(D)$ over the field \mathbf{F} remain true for $L(D)_{\text{rat.}}$ over the field \mathbf{F}_q .

From now on we refer to $L(D)_{\text{rat.}}$ as $L(D)$.

Remark 3.2.8 Let D be a divisor with $D = \sum n_i Q_i$ where $n_i \in \mathbf{Z}$ and f be a rational function in $L(D)$ then when $n_i < 0$ f has **zero** at Q_i of multiplicity at least $|n_i|$, and when $n_i > 0$ f has **pole** at Q_i with order at most n_i .

Let X be an absolutely irreducible non-singular projective curve over the field \mathbf{F}_q .

Let P_1, P_2, \dots, P_n be rational points on X and D be the divisor $D := P_1 + P_2 + \dots + P_n$. Furthermore let G be some other divisor on X that has support **disjoint** from D and also:

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

Definition 3.2.9 The linear code $C(D, G)$ of length n over the field \mathbf{F}_q is the image of the linear map

$$\alpha : L(G) \rightarrow \mathbf{F}_q^n$$

defined by $\alpha(f) = (f(P_1), f(P_2), \dots, f(P_n))$. Codes of this kind are called **geometric Reed-Solomon codes**.

The following theorem holds

Theorem 3.2.10 The code $C(D, G)$ has dimension $\kappa = \deg(G) - g + 1$, where g is the genus of the curve X , and minimum distance

$$d \geq n - \deg(G)$$

Proof $f \in L(G)$ implies that $\langle f \rangle + G \geq 0$.
If f belongs to the kernel of α i.e.

$$\alpha(f) = (f(P_1), f(P_2), \dots, f(P_n)) = (0, 0, \dots, 0)$$

which means that f has zeros on the points P_1, P_2, \dots, P_n of multiplicity at least one. Since $D = P_1 + P_2 + \dots + P_n$ we get that $-D = -P_1 - P_2 - \dots - P_n$. $-D$ has poles on the points P_1, P_2, \dots, P_n of order exactly one. So, $\langle f \rangle + G - D \geq 0$. This implies that $f \in L(G - D)$.

The degree of $-D$ is $\deg(-D) = -n$ and so the degree of the divisor $G - D$ is $\deg(G - D) = \deg G + \deg(-D) < n - n = 0$. Due to a known theorem of the algebraic curves theory when the degree of a divisor Δ is less than zero, then the dimension of the space $L(\Delta)$ is $l(\Delta) = 0$. (See [Ho], theorem 2.37, page 886). Thus, $L(G - D) = \{0\}$, so

$$f \in \ker \alpha \Leftrightarrow f = 0$$

hence, we get that the map α is one-to-one.

From the assumption: $2g - 2 < \deg(G) < n$. It occurs from the Riemann - Roch theorem (see [Ho], theorem 2.55, page 890) that if the degree of a divisor Δ , is $\deg(\Delta) > 2g - 2$ then it is of dimension $l(\Delta) = \deg(\Delta) - g + 1$ (see [Ho], corollary 2.58, page 890). Thus, $\kappa = l(G) = \deg(G) - g + 1$.

We prove now that the minimum distance of the code is more or equal to $n - \deg(G)$.

If $\alpha(f)$, $f \in L(G)$, has weight d then it is non zero in exactly d points from P_1, P_2, \dots, P_n . So it is zero in exactly $n - d$ points from P_1, P_2, \dots, P_n . Then $f(P_{i_1}) = f(P_{i_2}) = \dots = f(P_{i_{n-d}}) = 0$. If E is the divisor $E = P_{i_1} + P_{i_2} + \dots + P_{i_{n-d}}$ we get $\langle f \rangle + G - E \geq 0$ i.e. $f \in L(G - E)$. Therefore, $\deg G + \deg \langle f \rangle - \deg E \geq 0$. Due to theorem 3.2.6 we get $\deg \langle f \rangle = 0$ so $\deg G - \deg E \geq 0$ $\hat{=}$ $\deg G - (n - d) \geq 0$. Hence $d \geq n - \deg G$ and the proof is complete.

It is already clear that we find some **good** codes. If we apply the theorem 3.2.10 for curves of 0, we see that we get MDS codes. In fact, the theorem says that

$$d \geq n - \kappa + 1$$

and from the Singleton bound we get $\kappa \leq n - d + 1$. So, $\kappa = n - d + 1$ and we have an MDS code.

More general, for curves of small genus we get codes close to the Singleton bound.

3. Examples of algebraic geometry codes

We already know that to construct good codes, we must construct long codes. To use the methods from algebraic geometry, it is necessary to consider algebraic curves with many rational points. The number of these points is a bound on the length of the code. A central problem in algebraic geometry is finding (upper) bounds for the number of rational points on an algebraic curve or, more general, on a variety. So, it is very useful the Hasse – Weil theorem (theorem 2.3.3). In the following examples we use the following bound which is an improvement of the Hasse – Weil bound by Serre.

Theorem 3.3.1 (Serre Bound) Let X be a curve of genus g defined over the field \mathbf{F}_q . If $N_q(X)$ denotes the number of its rational points, then

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

(For the proof see [St], chapter 3, theorem 3.1, page 180)

Let's now state two examples of algebraic geometry codes.

Example 3.3.2 Let K_3 be the Klein quartic over the field \mathbf{F}_8

$$K_3: X^3Y + Y^3Z + Z^3X = 0$$

The curve is non-singular and has genus 3 (see definition 1.3.9). From the Serre bound (theorem 3.3.1) we have that the curve can have at most $\lfloor 2\sqrt{8} \cdot 3 \rfloor + (8 + 1) = 15 + 9 = 24$ rational points.

We show that has exactly 24 rational points. The field \mathbf{F}_8 is a simple extension of degree 3 of \mathbf{F}_2 and it is of the form $\mathbf{F}_2(\xi)$ where $\xi^3 = \xi + 1$. We will study the rational points of the curve X over the fields \mathbf{F}_2 and \mathbf{F}_8 . The rational points of the curve over the field are $[1, 0, 0]$, $[0, 1, 0]$, $[0, 0, 1]$. If we consider the curve over \mathbf{F}_8 a rational point $[x, y, z]$ has a coordinate zero if and only if it is one of the points over \mathbf{F}_2 . If $xyz \neq 0$, we can take $z = 1$. The multiplicative group \mathbf{F}_8^* is cyclic of order 7 generated from the element ξ . Because $y \neq 0$ we take $y = \xi^i$ ($0 \leq i \leq 6$). We can write that

$x = \xi^{3i} \eta$. Substitution in the equation K_3 gives us

$$\xi^{9i} \eta^3 \xi^i + \xi^{3i} + \xi^{3i} \eta = 0$$

i.e.

$$\xi^{7i} \eta^3 + 1 + \eta = 0$$

i.e.

$$\eta^3 + \eta + 1 = 0$$

Thus, η is a solution of the equation $X^3 + X + 1 = 0$. This means that η is one of the elements ξ, ξ^2, ξ^4 .

So we find that K_3 has totally $3 + 7 \cdot 3 = 24$ rational points.

Let $Q = (0, 0, 1)$ and D is the sum of the remaining 23 rational points. Let $G = 10Q$. From theorem 3.2.10 we get that the code $C(D, G)$ we constructed is dimension equal to $\kappa = \deg(G) - g + 1 = 10 - 3 + 1 = 8$ and minimum distance $d \geq 23 - 10 = 13$. Thus, it is a $(23, 8, 13)$ -code. We now concatenate this code with the $(4, 3, 2)$ -single parity check code as follows: The symbols in codewords of $C(D, G)$ are elements of \mathbf{F}_8 which we interpret as column vectors of length 3 over \mathbf{F}_2 and then we adjoin the parity check. The resulting code C is a binary $(92, 24, d)$ -code with $d \geq 13 \cdot 2 = 26$ (See remark 3.1.34). The punctured code of C is a $(91, 24, d - 1)$ -code. For $d = 26$ the punctured $(91, 24, d - 1)$ -code set a new world record for codes with $n = 91$ and $d = 25$. (See [Ho], chapter 10, paragraph 2.8, example 2.75).

Example 3.3.3 We consider the field $\mathbf{F}_4 = \{0, 1, a, \bar{a}\}$ where $a^2 = a + 1 = \bar{a}$ and $a^3 = 1$. The field's characteristic is $\text{ch}\mathbf{F}_4 = 2$. We consider the curve X over \mathbf{F}_4 given by the equation:

$$X: X^2Y + aY^2Z + \bar{a}Z^2X = 0$$

The curve X is non-singular because if there exists a singular point $P = [x, y, z]$ on the curve we must have: $\frac{\partial}{\partial X}X|_{P=(x,y,z)} = \frac{\partial}{\partial Y}X|_{P=(x,y,z)} = \frac{\partial}{\partial Z}X|_{P=(x,y,z)} = 0$ i.e. $2xy + \bar{a}z^2 = 2ayz + x^2 = 2\bar{a}z + ay^2 = 0$ i.e. $\bar{a}z^2 = x^2 = ay^2 = 0$ i.e. $x = y = z = 0$, which is a contradiction.

Thus, X is of genus 1 and the Serre bound implies that $N_q(X) \leq (4 + 1) + 1 [2\sqrt{4}] = 9$. So the curve has at most 9 rational.

As we can see from the following table the curve has exactly 9 rational points:

	P_1	P_2	P_3	P_4	P_5	P_6	Q_1	Q_2	Q_3
x	1	0	0	1	1	1	a	1	1
y	0	1	0	A	\bar{a}	1	1	A	1
z	0	0	1	\bar{a}	a	1	1	1	a

Let $D = P_1 + P_2 + \dots + P_6$ and $G = 2Q_1 + Q_2$.

The degree of the divisor G is $\deg(G) = 2 + 1 > 2g - 2 = 2 \cdot 1 - 2 = 0$. So, it is an immediate corollary from the Riemann - Roch theorem (see [Ho], theorem 2.55, page 890) that

$$l(G) := \dim L(G) = \deg(G) - g + 1 = 3 - 1 + 1 = 3$$

We claim that the rational functions

$$\begin{aligned} f_1(X, Y, Z) &:= \frac{X}{X + Y + \bar{a}Z}, \\ f_2(X, Y, Z) &:= \frac{Y}{X + Y + \bar{a}Z}, \\ f_3(X, Y, Z) &:= \frac{\bar{a}Z}{X + Y + \bar{a}Z} \end{aligned}$$

form a basis of $L(G)$.

Due to remark 3.2.8, in order to prove that $f_1(X, Y, Z), f_2(X, Y, Z), f_3(X, Y, Z) \in L(G)$ we just have to prove that in the points Q_1 and Q_2 have poles of order at most 2 and 1 respectively. We can easily check that the points Q_1 and Q_2 are poles of f_1, f_2, f_3 . We just note that for these functions the numerator, in these points, is not zero and for the denominator are each of f_1, f_2, f_3 it holds:

$$\begin{aligned} (X + Y + \bar{a}Z) \Big|_{Q_1=(a,1,1)} &= a + 1 + \bar{a} = a^2 + a + 1 = 0 \\ (X + Y + \bar{a}Z) \Big|_{Q_2=(1,a,1)} &= 1 + a + \bar{a} = a^2 + a + 1 = 0 \end{aligned}$$

The points Q_1 and Q_2 are intersection points of the line $\varepsilon : X + Y + \bar{a}Z = 0$ and the cubic curve X because ε intersects X in Q_2 and the tangent of the curve X is

$$\frac{\partial}{\partial X} X \Big|_{Q_1=(a,1,1)} X + \frac{\partial}{\partial Y} X \Big|_{Q_1=(a,1,1)} Y + \frac{\partial}{\partial Z} X \Big|_{Q_1=(a,1,1)} Z = 0.$$

Thus, $a^2 X + a^2 Y + a Z = 0 \hat{=} a^3 X + a^3 Y + a^2 Z = 0 \hat{=} X + Y + \bar{a}Z = 0$

By the Bezout's theorem (theorem 1.3.19) the intersection point of the line ε and the curve X are three and the intersection multiplicity is exactly 2 in Q_1 and 1 in Q_2 .

So we conclude that $f_1(X, Y, Z), f_2(X, Y, Z), f_3(X, Y, Z) \in L(G)$. Because they are linearly independent they are a basis of $L(G)$.

By the theorem 3.2.10 we have that the code $C(D, G)$ of length 6 has minimum distance $d \geq 6 - 3 = 3$. The inequality $\kappa \leq n - d + 1$ from the Singleton bound (theorem 3.1.18) implies that $3 \leq 6 - d + 1 \hat{=} d \leq 4$. The code is equivalent to the namely Hexacode, they have the same generator matrix. By using methods from the finite geometry it has been proven that the minimum distance of the Hexacode is $d = 4$. So, $C(D, G)$ has also minimum distance $d = 4$ i.e. $\kappa = n - d + 4$, which implies that $C(D, G)$ is an MDS code. (For the Hexacode see [Pe]).

Bibliography

- [A1] J. A. Antoniadis, Elliptic curves (Mordell's theorem), "Prometheus" (EPEAEK program), Heraklio 1999
- [A2] J. A. Antoniadis, Applied Algebra, "Prometheus" (EPEAEK program), Heraklio 2000
- [Ch1] J. S. Chahal, Manin's Proof of the Hasse Inequality Revisited, Nieuw Archief voor Wiskunde, pp. 219 – 232, Vierde serie Deel 13 No 2, juli 1995
- [Ch2] J. S. Chahal, Topics in Number Theory, Plenum Press, New York 1988
- [La] K. Lakki, Algebra, Thessaloniki 1993
- [MS] F. J. MacWilliams, N. J. A. Sloane, The Theory of Error-Correcting Code, North-Holland Mathematical Library, Sixth printing: 1988
- [Ho] Tom Hoholdt, J. H. van Lint, Ruud Pellikaan (authors), V. S. Pless, W. C. Huffman and R. A. Brualdi (editors), Handbook of Coding Theory, pp. 871-961, Elsevier Science, Amsterdam, 1998
- [Le] Franz Lemmermeyer, Elliptische Kurven I, available on the web: <http://www.rzuser.uni-heidelberg.de/~hb3/ellc.html>
- [LP] Rudolf Lidl - Gunter Pilz, Applied Abstract Algebra, Springer-Verlag 1998
- [Pe] Mario De Boer, Ruud Pellikaan, Gröbner bases for error-correcting codes and their decoding, Some tapes of computer algebra, Springer, Berlin 1999
- [S1] Joseph H. Silverman, John Tate, Rational Points on Elliptic Curves, Springer-Verlag 1992
- [S2] Joseph H. Silverman, The Arithmetic of Elliptic Curves, Springer-Verlag 1986
- [St] Henning Stichtenoth, Algebraic Function Field and Codes, Springer-Verlag 1993
- [W] R. J. Walker, Algebraic Curves, Dover, New York 1962

Index

Addition of points

co-linear over an elliptic curve, 20

Addition formulas for elliptic curves over \mathbf{Q} , 20

Addition formulas for elliptic curves over \mathbf{F}_q , 23

Addition formulas for elliptic curves over $\mathbf{F}_q[x]$, 42, 54, 55

Alphabet, 61

Information rate, 63

Distance

minimum, 62

Hamming, 61

Decoding

nearest neighbor, 61

Affine plane, 5

Affine curve, 7, 71

See also Curve

Degree

of an algebraic curve, 7

of a divisor, 71

of a principle divisor of a rational function, 72

of a homogeneous polynomial, 7

of a point's multiplicity, 11, 71, 72

Basis of a vector space, 77

Basic identity, 44, 48, 51, 53

Hamming weight, 62

Genus of a non-singular curve, 12, 38, 39

Vector

error, 61

received, 61

Divisor of a curve, 71

principle, 71

rational, 71

effective, 71

Vector space

- of a divisor, 71
- finite dimensional, 71, 72
- of a rational divisor, 72

Artin's conjecture, 38

Riemann hypothesis, 39

- for algebraic function fields of one variable with coefficients from \mathbf{F}_q , 38
- for elliptic curves, 39

Frobenius endomorphism, 37

Elliptic curve, 20, 27
See also Cubic Curve

Line

- at infinity, 5, 6, 9
- tangent, 10
- defined by two points, 9
- rational, 19
- in the projective plane, 6, 10

Theorem

- Reduction modulo p , 30
- Bezout, 13, 77
- Gauss, 38
- Hasse, 39, 40, 41, 46
- Lutz-Nagell, 20, 31, 33
- Mazur, 35
- Mordell, 20
- Riemann-Roch, 73, 76
- Weil (Hasse-Weil), 38

Isomorphism

- group, 30
- elliptic curve, 41, 42

Curve

- See also* Cubic curve, Elliptic curve
- affine, 7, 71
- irreducible plane algebraic, 7
- absolutely irreducible, 71
- genus of, 12
- plane algebraic, 7, 12, 13
- singular, 11
- cubic, 8
- non-singular, 11
- over a field K , 24
- over \mathbf{F}_4 , 76
- over \mathbf{F}_8 , 75
- coefficient, 8

order, 10
quartic, 8, 75

Cubic curve

See also Curve, Elliptic Curve
γενικευμένη μορφή Weierstrass, 17
discriminant of, 18, 23, 27, 61
singular point, 13
singular, 15, 16
normal form, 17
Weierstrass form, 18

Code, 61

error-detecting, 63, 64
binary, 76
dual, 68
geometric Reed-Solomon, 72
linear, 61, 66, 67
maximum distance separable, 64
error-correcting, 63, 64
repetition, 65
equivalent, 68
group, 65
ορθογώνιος, 68
punctured, 76
systematic linear, 65
perfect, 64
Golay, 67
MDS, 64, 73

Codeword, 61

Conic, 8, 13

irreducible, 15, 16
singular point, 13

Error (or error vector), 61

Null space, 65

Message,

received, 61
k-message, 61

Group

abelian, 19
cyclic, 25, 75
of rational points of an elliptic curve, 20, 24, 42
of rational points of finite order on an elliptic, 28
finite generated, 20
finite, 25

- of points of an elliptic curve over \mathbf{F}_3 , 31
- of points of an elliptic curve over \mathbf{F}_5 , 24, 30, 32, 34
- of points of an elliptic curve over \mathbf{F}_7 , 30, 32, 34
- of points of an elliptic curve over \mathbf{F}_p , 32

Group homomorphism, 28

Matrix

- generator, 66
- parity-check, 65
- canonical basic, 66
- encoding, 66

Projective plane, 5

Πρόσθεση σημείων, 19

See also Addition of points

Σειρά Taylor, 10

Point

- See also* Πρόσθεση σημείων, Άθροισμα σημείων
- Single, double, etc. 11
- at infinity, 5
- non-singular, 11, 16, 18
- singular, 11, 12
- ιδιάζον κυβικής καμπύλης, 13
- ιδιάζον κωνικής τομής, 13
- καμπής, 15, 16
- καμπύλης, 10, 19, 23, 75, 76
- πεπερασμένης τάξης, 27
- rational over \mathbf{Q} , 19
- rational over \mathbf{F}_q , 23, 24
- τομής δυο καμπυλών, 13
- τομής δύο κυβικών τομών, 13
- τομής δύο κωνικών τομών, 13
- τομής κυβικής καμπύλης με ευθεία, 19
- Frobenius, 71

Control symbols, 61

Legendre symbol, 45, 46

Co-linear points, 20, 29

Sphere about a codeword, 63

Order

- of a curve, 10
- of a group of rational points, 38, 39
- of a pole, 71, 72

of a rational point on an elliptic curve, 20

Intersection

See also conic

of a line with a curve, 10

Support of a divisor, 71

disjoint, 72

Bound

Gilbert-Varshamov, 69

Hamming, 64

Hasse – Weil, 38, 75

Plotkin, 69

Serre, 75, 76

Singleton, 64, 73, 76

Bombieri, 39

concatened codes, 69

discrete sphere packing problem, 64

Hexacode, 77

Lemmermeyer, 47

Manin, 39

Roquette, 39, 47

Stepanov, 39

Weil, 38

Twist, 41

