Elliptic Curves and Mathematical Cryptography

Department of Mathematics, Linköping University

Tove Söneby

LiTH-MAT-EX–2025/16–SE

Credits: **16 hp**

Level: **G2**

Supervisor: **Milagros Izquierdo**,
Department of Mathematics, Linköping University

Examiner: **Göran Bergqvist**,
Department of Mathematics, Linköping University

Linköping: **December 2025**

Abstract

Elliptic curves are the basis for Elliptic Curve Cryptography (ECC) that utilise points on a curve for its public and private keys. A special kind of addition, point addition, adds two points on the curve to get a third point on the same curve. In order for this to be cryptographically useful the addition takes place over a finite field. Elliptic Curve Cryptography has various uses, maybe most notably in the use of blockchains. ECC has the advantage of offering higher security for shorter keys compared to for example RSA. To motivate the existence of this point relation we explore algebraic curves, points and lines at infinity as well as intersections between curves.

Keywords:

Elliptic curves, point addition, cryptography, intersection of curves, finite fields

URL for electronic version:

`http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-220227`

Sammanfattning

Elliptiska kurvor ligger till grund för Elliptisk-kurv-kryptografi, ECC (Elliptic Curve Cryptography), som använder sig av punkter på en kurva till sina publika och privata nycklar. En särskild typ av addition, punktaddition, adderar två punkter på kurvan för att få en tredje punkt på samma kurva. För att det ska vara kryptografiskt användbart sker additionen över ändliga kroppar. Elliptisk-kurv-kryptografi har flera tillämpningar, där det kanske mest uppmärksammade är i blockkedjor. ECC har fördelen av att erbjuda högre säkerhet med kortare nycklar jämfört med exempelvis RSA. För att säkerställa existensen av punktadditionen undersöker vi algebraiska kurvor, oändlighetspunkter och -linjer och skärningar mellan kurvor.

Nyckelord:

Elliptiska kurvor, punktaddition, kryptografi, skärningar mellan kurvor, ändliga kroppar

URL för elektronisk version:

`http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-220227`

Acknowledgements

First of all I'd like to thank my supervisor Milagros Izquierdo for her guidence and patience during my work with this essay. She is always answering any questions I might have, regardless of day, time of day or where in the world she might be. Thank you to all of my friends and family who has supported me throughout this process, with an extra, extra big thanks to my partner.

Nomenclature

G	group
R	ring
D	domain
$D[x]$	set of all polynomials in x over D
$D[x_1, \ldots, x_r]$	set of all polynomials in $x_1, \ldots, x_r$ over D
$\mathbb{Z}$	the integers
$\mathbb{Q}$	the rational numbers
$\mathbb{R}$	the real numbers
$\mathbb{C}$	the complex numbers
$\mathbb{Z}/m\mathbb{Z}$	ring of integers modulo m
$(\mathbb{Z}/m\mathbb{Z})^*$	group of units modulo m
K	field
$\overline{K}$	algebraic closure of K
$\mathbb{F}_p, \mathbb{F}_{p^d}$	finite field with p, resp. p^d elements, p prime
$(\mathbb{F}_p)^*$	group of units modulo p, p prime
f'	derivative of f
$\frac{\partial^r f}{\partial x^r}$	r:th partial derivative of f with regard to x
$\mathbb{P}^n(K)$	n-dimensional projective space over the field K
$\oplus$	point addition
∇f	gradient for the function f
E	elliptic curve
$E(\mathbb{F}_p)$	elliptic curve over the finite field $\mathbb{F}_p$
$\mathcal{O}$	point at infinity

Contents

Chapter 1

Introduction

Encrypted messages most likely predate Julius Caesar (100 B.C) and the cipher that share his name. Without a doubt, cryptography has played a great role throughout history, from secret messages written by hand on paper to the mechanical Enigma machine. Today, when so much communication is done digitally, cryptography is done on computers, by computers and computers try to break the cryptographic algorithms. Probably the most known modern cryptosystem is the RSA cryptosystem which utilises congruence calculation with large prime numbers. Less known to the broader public is a category of cryptosystems based on points on certain curves called *Elliptic Curve Cryptography*, or ECC for short. This kind of cryptosystem utilises a relation between points on an elliptic curve. A particular addition, *point addition*, connects two given points on the elliptic curve with a third point on the same curve. By letting these calculations be cyclical within a finite system of numbers, calculations within these systems can be applied to construct algorithms similar to the congruence calculations used in RSA. This point addition was first applied to an elliptic curve called a *Weierstrass curve*. In 2007, Harold Edwards presented a normal form for the Weierstrass curve, called an *Edwards curve*, onto which the same type of point addition could be applied. The Edwards curve has some properties that makes it more desirable to use compared to the Weierstrass curve. The point addition for a Weierstrass curve is performed differently depending on the points being added, while the point addition for an Edwards curve is the same regardless of choice of points.

But we are getting ahead of ourselves. In order for any of these calculations to be made possible there are several mathematical concepts that need to be present. This essay will cover what kind of structures are needed in order for these curves and this addition to work. It will also cover how calculations

work in these structures, what curves are and how they relate to each other. Chapter 2 will cover the foundational theory needed for elliptic curves. We will start with the underlying structures, the general *field* and the *finite field* and describe how operations are done on these structures. We then proceed to the concept of *projective spaces* and assigning fixed coordinates to the infinity. Then we will move on to curves and how one can determine intersections between two curves. In chapter 3 we will introduce the elliptic curves. We will start with the Weierstrass curve and its point addition, both on a general field and a finite field. Then we will introduce the Edwards curve and its point addition. Some effort will be put towards demonstrating the relation between the Weierstrass and Edwards curves. Finally, in chapter 4 we will look at some of the practical applications that utilises elliptic curves. Both the cryptosystems in themselves as well as a highly current topic that uses a particular elliptic curve algorithm.

Chapter 2

The Underlying Theory

This chapter will cover all the different topics that lay the foundation for elliptic curves. First, in section 2.1, we will look at the structures which constitute the basis for all of the other topics in this essay, groups, rings, domains and fields. In section 2.2 we will explore a particular kind of fields with a finite number of elements, finite fields. After that we will cover resultants and discriminants, which are used to determine common factors of polynomials. Then, projective spaces which let us view the infinity as points or lines. The last two sections of this chapter will cover curves and how to determine where two curves intersect. The definitions and theorems in this chapter are sourced from [1], [2], [3] and [6]. Some are reproduced as they appear in the source material, others are shortened or reworked in order to fit this essay better.

2.1 Groups, Rings, Domains and Fields

In this section we will present a number of structures consisting of a set, one or two binary operator and a series of axioms. A *Group* is the most general structure we will encounter in this essay and constitutes the basis for the other structures. Each subsequent structure will add additional conditions to that of groups. We will mainly be interested in *fields* and their properties for they are the foundation on which this essay rests upon. The sources used in this section are [1], [2], [3] and [6].

Definition 2.1.1 (Group)**.** A *group*, denoted $(G, \cdot)$, consists of a set G and a binary operator "$\cdot$" defined on G for which the following three conditions are satisfied:

1. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in G$ (*associative*).

2. There is $1 \in G$ such that $1 \cdot a = a \cdot 1$ for all $a \in G$ (*identity*).

3. Given $a \in G$, there is $b \in G$ such that $a \cdot b = b \cdot a = 1$ (*inverses*).

If also $a \cdot b = b \cdot a$ for all $a, b \in G$ (*commutative*), $(G, \cdot)$ is called a *abelian group* or a *commutative group*.

An example of an abelian group is the integers with the standard additive operator $+$, denoted $(\mathbb{Z}, +)$. That $(\mathbb{Z}, +)$ is associative and commutative is easy to see. For the identity and inverses criteria, the identity is the integer 0 and the inverses are the opposite sign of any given integer. Moreover, the identity for abelian groups are often called 0.

Definition 2.1.2. Let G be a group. A subset $H \subset G$ is called a *subgroup* if it satisfies:

1. $1 \in H$.

2. If $a, b \in H$, then $a \cdot b \in H$.

3. If $a \in H$, then $a^{-1} \in H$.

Example 2.1.3. The group of even integers, denoted $(2\mathbb{Z}, +)$, is a subgroup of the integers $(\mathbb{Z}, +)$. The identity $0 \in (2\mathbb{Z}, +)$, any even number added to another is also a even number and the negative even numbers are inverses to the positive even numbers and vice versa.

Definition 2.1.4 (Ring)**.** A *ring*, denoted $(R, +, \cdot)$, consists of a set R and two binary operators, a additive operator "$+$" and a multiplicative operator "$\cdot$" where

1. $(R, +)$ is an abelian group.

2. The multiplicative operator, $\cdot$, is associative.

3. The multiplicative operator, $\cdot$, distributes over the additive operator, $+$. That is, $a \cdot (b + c) = a \cdot b + a \cdot c$.

If there exists a multiplicative identity, $(R, +, \cdot)$ is said to be a *ring with identity*. Moreover, if the multiplicative operator is commutative it is said to be a *commutative ring*.

An example of a non-commutative ring with identity are the real $n \times n$-matrices with $n \geq 2$ with the standard matrix addition and multiplication. The multiplicative identity element is the identity matrix and $A + B = B + A$ but $AB \neq BA$ in general.

Remark. In this essay only commutative rings with identity will be considered.

We will take a quick look at a specific ring, *the ring of integers*. This is interesting mainly for comparative reasons later on. The ring of integers functions similarly to that of finite fields, section 2.2, but there are some distinct differences regarding inverses. For now we will be content with the definition and an example.

Definition 2.1.5. *The ring of integers modulo* m, denoted $(\mathbb{Z}/m\mathbb{Z}, +, \cdot)$, is the set $\mathbb{Z}/m\mathbb{Z} = \{0, 1, \ldots, m-1\}$ with addition and multiplication calculated (mod m).

Example 2.1.6. Let us look at the ring of integers modulo 4, $(\mathbb{Z}/4\mathbb{Z}) = \{0, 1, 2, 3\}$, with binary operators $+$ and $\cdot$ that are calculated (mod 4). Some examples of addition and multiplication done in this ring are

$$\begin{aligned} 0+1&=1 & \qquad 1\cdot 2&=2\\ 1+2&=3 & \qquad 2\cdot 2&=4=0\\ 2+3&=5=1 & \qquad 2\cdot 3&=6=2 \end{aligned}$$

Take particular notice at the multiplication $2 \cdot 2 = 0$. This implies that 2 is a zero divisor (which is not permitted in domains or fields, see 2.1.7 resp. 2.1.10). The complete addition and multiplication tables are listed in table 2.1.

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	1	2	3
1	1	2	3
2	2	0	2
3	3	2	1

Table 2.1: Addition and multiplication tables for the ring of integers modulo 4, $\mathbb{Z}/4\mathbb{Z}$.

Definition 2.1.7 (Domain)**.** A *integral domain* or *domain*, denoted D, is a commutative ring with identity where whenever $a, b \in D$ and $a \cdot b = 0$, then either $a = 0$ or $b = 0$.

The most familiar example of a domain is the integers, $\mathbb{Z}$, with standard addition and multiplication. Notice that both domains and rings have additive inverses but neither are required to have multiplicative inverses. This is an important distinction from a field, which we shall define shortly.

Theorem 2.1.8. *Let D be a domain. If for an element $a \neq 0$, $a \in D$, there exists a positive integer m such that $ma = 0$, then $mb = 0$ for every $b \in D$. The smallest such integer n is called the* characteristic *of D, denoted* $\text{char}(D) = n$. *If no such integer exists, D is said to have characteristic zero.*

Some examples of domains with characteristic zero are $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$. Examples of domains with non-zero characteristic is the finite fields $\mathbb{F}_p$ and $\mathbb{F}_{p^d}$ both with characteristic p, as we will see in section 2.2.

Definition 2.1.9. Let D be a domain. The elements of D are called *constants*. Corresponding to any ordered finite set $a_0, a_1, \ldots, a_n$ of constants we form the expression $a_0x^0 + a_1x^1 + \ldots + a_nx^n$, and call it a *polynomial in x over D*. The symbol x is called a *variable*. It is introduced to facilitate computation and is not an element of D. The set of all polynomials in x over D is denoted $D[x]$. $D[x]$ is also a domain.

Definition 2.1.10 (Field)**.** A *field*, denoted $(K, +, \cdot)$, consists of a set K and two binary operators, an additive operator "+" and a multiplicative operator "·" where:

1. $(K, +)$ is an abelian group.
2. $(K^* = K \setminus \{0\}, \cdot)$ is an abelian group.
3. The multiplicative operator, $\cdot$, distributes over the additive operator $+$. That is, $a \cdot (b + c) = a \cdot b + a \cdot c$.

Some examples of fields are the sets $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ with the standard addition and multiplication. However, $\mathbb{Z}$ is not a field because, generally, elements in $\mathbb{Z}$ do not have a multipliative inverse. The only exceptions are the elements 1 and -1, which are their own multiplicative inverses. Another example of a field is $\mathbb{F}_p$, p prime, see section 2.2.

Definition 2.1.11. Let K be a field. A subset $L \subset K$ is called a *subfield* if it satisfies:

1. The additive and the multiplicative identities are in L, i.e. $0, 1 \in L$.
2. L is closed under the binary operators.
3. If $a \in L$, where $a \neq 0$, then $-a \in L$ and $a^{-1} \in L$.

Definition 2.1.12. The *prime subfield* of a field K is the subfield of K generated by the multiplicative identity 1 of K. It is (isomorphic to) either $\mathbb{Q}$ (if $\text{char}(K) = 0$) or $\mathbb{F}_p$ (if $\text{char}(K) = p$).

Definition 2.1.13. If L is a field containing the field K, then L is said to be an *extension field* (or simply an *extension*) of K, denoted L/K. In particular, every field K is an extension of its prime subfield.

Definition 2.1.14. The element $\alpha \in L$ is said to be *algebraic* over K if α is a root of any non-zero polynomial $f(x) \in K[x]$. If α is not algebraic over K (i.e. is not the root of any non-zero polynomial with coefficients in K) then α is said to be *transcendental* over K. The extension L/K is said to be *algebraic* if every element of L is algebraic over K.

Definition 2.1.15. The extension field L of K is called a *splitting field* for the polynomial $f(x) \in K[x]$ if $f(x)$ factors completely into linear factors (or *splits completely*) in $L[x]$ and $f(x)$ does not factor completely into linear factor over any proper subfield of L containing K.

Definition 2.1.16. The field $\overline{K}$ is called an *algebraic closure* of K if $\overline{K}$ is algebraic over K and every polynomial $f(x) \in K[x]$ splits completely over $\overline{K}$ (so that $\overline{K}$ can be said to contain all the elements algebraic over K).

Remark. The practical use of the algebraic closure is that calculations performed in a field K can be viewed as being performed in its algebraic closure $\overline{K}$.

A lot of properties for domains and fields, such as arithmetic for sums, products and powers etc., function as one would expect and are familiar with from other areas of mathematics and will not be discussed in this essay. If the reader is interested in the specifics and wish to learn more, [6] makes a thorough exposition.

The definitions and theorems about domains also apply to fields. All fields are domains and so all properties of domains also apply to fields (however, the converse is not true). Keep this in mind since fields are the basis for most of the other concepts of this essay.

Definition 2.1.17. Let D be a domain and let $a \in D$.

1. An element $e \in D$ which has a multiplicative inverse, e^{-1}, is called a *unit*. If e is a unit, ea is said to be an *associate* of a.

2. The units of D form a commutative group under multiplication.

3. Any element of D is divisible by its associates and by units. An element is called *irreducible* if it is divisible only by its associates and units.

As previously mentioned, in the case of $\mathbb{Z}$ the only multiplicative inverses that exist, i.e. its units, are 1 and -1. Therefore, the positive integers are the negative integers associates and vice versa. And familiarly, the prime numbers and their negative associates are the irreducible elements of $\mathbb{Z}$.

Definition 2.1.18. The set $(\mathbb{Z}/m\mathbb{Z})^*$ is called the *group of units modulo* m, where

$$(\mathbb{Z}/m\mathbb{Z})^* = \{a \in \mathbb{Z}/m\mathbb{Z} : a \text{ has a inverse modulo } m\} = \\ = \{a \in \mathbb{Z}/m\mathbb{Z} : \gcd(a, m) = 1\}.$$

Example 2.1.19. Let $\mathbb{Z}/6\mathbb{Z} = \{0, \ldots, 5\}$ and $\mathbb{Z}/9\mathbb{Z} = \{0, \ldots, 8\}$. Then $(\mathbb{Z}/6\mathbb{Z})^* = \{1, 5\}$ and $(\mathbb{Z}/9\mathbb{Z})^* = \{1, 2, 4, 5, 7, 8\}$. The multiplication tables for $(\mathbb{Z}/6\mathbb{Z})^*$ and $(\mathbb{Z}/9\mathbb{Z})^*$ are listed in table 2.2.

$(\mathbb{Z}/6\mathbb{Z})^*, \cdot$	1	5
1	1	5
5	5	1

$(\mathbb{Z}/9\mathbb{Z})^*, \cdot$	1	2	4	5	7	8
1	1	2	4	5	7	8
2	2	4	8	1	5	7
4	4	8	7	2	1	5
5	5	1	2	7	8	4
7	7	5	1	8	4	2
8	8	7	5	4	2	1

Table 2.2: Multiplication tables for $(\mathbb{Z}/6\mathbb{Z})^*$ and $(\mathbb{Z}/9\mathbb{Z})^*$.

Definition 2.1.20. D is said to have a *unique factorisation* if

1. For every element $a \neq 0$ of D there exists a finite set of irreducible elements $a_1, \ldots, a_r$ whose product is a.

2. If $a_1 a_2 \cdot \ldots \cdot a_r = b_1 b_2 \cdot \ldots \cdot b_s$, the a_i, b_i being irreducible non-units, then $r = s$ and the factors can be arranged so that a_i is associate to b_i.

Theorem 2.1.21. *If D has unique factorisation so has $D[x_1, \ldots, x_r]$.*

Example 2.1.22. Let $D = \mathbb{Z}$. In $D[x] = \mathbb{Z}[x]$ we have

$$6x^3 + 23x^2 + 25x^1 + 6x^0 = (2x + 3)(x + 2)(3x + 1)$$

as an example of a unique factorisation in one variable.
In $D[x, y] = \mathbb{Z}[x, y]$ we have

$$2x^2y + 4x^2 + 6xy^2 + 15xy + 6x + 9y^2 + 18y = (2x + 3)(y + 2)(3y + x)$$

as an example of unique factorisation in two variables.

In $D[x,y] = \mathbb{Q}[x,y]$ we have

$$\begin{aligned} 2x^2y + \frac{3x^2}{5} + 6xy^2 + \frac{37xy}{15} + \frac{x}{5} + 2y^2 + \frac{3y}{5} = \\ = \left(x + \frac{1}{3}\right)\left(2y + \frac{3}{5}\right)(x + 3y) = \\ = \frac{1}{15}(3x+1)(10y+3)(x+3y) \end{aligned}$$

as another example of unique factorisation in two variables.

Remark. In the case of polynomials in $\mathbb{Q}$, remember that all non-zero elements in $\mathbb{Q}$ have a multiplicative inverse and are therefore units. Thus, the two variants of polynomials in $\mathbb{Q}[x,y]$ is not a violation of unique factorisation.

Theorem 2.1.23. *If $f(x_1,\ldots,x_r) \in D[x_1,\ldots,x_r]$, and if $f(a_1,\ldots,a_r) = 0$ for all choices of $a_1,\ldots,a_r$, from any fixed infinite subset of D, then $f(x_1,\ldots,x_r) = 0$.*

Definition 2.1.24. A domain D is said to be *algebraically closed* if for every non-constant $f \in D[x]$ there is an element $a \in D$ such that $f(a) = 0$.

Neither $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{R}$ are algebraically closed because $x^2+1=0$ has no solution in either of these domains. However, $\mathbb{C}$ is algebraically closed since $x^2+1=0$ has solutions in $\mathbb{C}$. Observe that an algebraically closed domain is necessarily a field, since $bx - 1 = 0$ has a root for any $b \neq 0$. That is, $b \in \mathbb{C} \setminus \{0\}$ has a multiplicative inverse.

Theorem 2.1.25. *If D is algebraically closed and if $f \in D[x]$ is of degree n, then there exists a unique set of n constants $a_1,\ldots,a_n$, not necessarily distinct, such that*

$$f(x) = a(x-a_1)\cdot\ldots\cdot(x-a_n),\ a \in D,\ a \neq 0.$$

This applies only to polynomials of one variable.

2.2 Finite Fields

Up till now our examples of fields have been those with an infinite number of elements such as $\mathbb{R}$ and $\mathbb{C}$. However, it is entirely possible to construct fields with a finite number of elements. *Finite fields* and their elements are what is used when creating cryptosystems based on elliptic curves, which we will see in chapters 3 and 4. The source used in this section is [3].

Definition 2.2.1. A *finite field* is a field with a finite number of elements.

When discussing finite fields we will generally use the notation $\mathbb{F}_p$ or $\mathbb{F}_{p^d}$ for fields with p respective p^d number of elements, rather than the more general notation K.

Theorem 2.2.2. *Let $\mathbb{F}_q$ be a finite field having $q = p^d$ elements, p prime. Then $\mathbb{F}_q$ has a primitive root, i.e. there is an element $g \in \mathbb{F}_q$ such that*

$$\mathbb{F}_q^* = \{1, g, g^2, g^3, \ldots, g^{q-2}\}.$$

Example 2.2.3. If p is prime, then the ring of integers modulo p, $\mathbb{Z}/p\mathbb{Z}$, is a finite field.

Observe that $\mathbb{Z}/p\mathbb{Z}$ and $\mathbb{F}_p$, p prime, actually are different notations for the same object. Similarly, $\mathbb{F}_p^*$ and $(\mathbb{Z}/p\mathbb{Z})^*$ also denote the same object.

Example 2.2.4. Let $\mathbb{F}$ be a field and let $\boldsymbol{m} \in \mathbb{F}[x]$ be an irreducible polynomial. Then the quotient ring $\mathbb{F}[x]/(\boldsymbol{m})$ is a field.

It is fairly straight forward to construct a finite field with p number of elements where p is prime. The simplest representation of $\mathbb{F}_p$ is $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$ which is, as previously mentioned, equal to $\mathbb{Z}/p\mathbb{Z}$. To construct a finite field with p^d elements is not as simple. Recall example 2.1.6 about the ring of integers modulo 4, $\mathbb{Z}/4\mathbb{Z}$. As seen in table 2.1, $2 \cdot 2 = 0$ which implies that 2 is a zero divisor which is a violation of the definition for domains, 2.1.7, which also applies to fields. Therefore, we can not use $\mathbb{Z}/m\mathbb{Z}$, where $m = p^d$, to construct a finite field with p^d elements. There are only finite fields with p or p^d, p prime, $d \in \mathbb{Z}_+$, number of elements. There will be no proof given as to why there is no finite fields with composite number of elements that are not prime powers. But as hinted above, a short explanation is that trying to construct a finite field with composite number of elements will yield several distinct elements as zero divisors. So, to construct a finite field with, for example, four elements we have to use a different method. We will utilize polynomials for this purpose.

Theorem 2.2.5. *Let $\mathbb{F}_p$, p prime, be a finite field.*

1. *For every $d \geq 1$ there exists an irreducible polynomial $\boldsymbol{m} \in \mathbb{F}_p[x]$ of degree d.*
2. *For every $d \geq 1$ there exists a finite field with p^d elements.*
3. *If $\mathbb{F}$ and $\mathbb{F}'$ are finite fields with the same number of elements, then there is a way to match the elements of $\mathbb{F}$ with the elements of $\mathbb{F}'$ so that the addition and multiplication tables of $\mathbb{F}$ and $\mathbb{F}'$ are the same. I.e. $\mathbb{F}$ and $\mathbb{F}'$ are isomorphic.*

Example 2.2.6. The polynomials of $\mathbb{F}_2[x]$ are 0, 1, x, $x+1$, x^2, x^2+1, x^2+x, x^2+x+1, etc. To find the irreducible polynomials, start with the polynomials known too be irreducible and eliminate all polynomials that they factor. So in this case the polynomials of degree 1, x and $x+1$, are the first irreducible polynomials in $\mathbb{F}_2[x]$. We have that

$$\begin{aligned} x \cdot x &= x^2 \\ x(x+1) &= x^2 + x \\ (x+1)(x+1) &= x^2 + 2x + 1 = x^2 + 1 \qquad \text{since } 2 = 0 \text{ in } \mathbb{F}_2 \end{aligned}$$

are the reducible polynomials of degree 2. Therefore the only irreducible polynomial in $\mathbb{F}_2[x]$ of degree 2 is $x^2 + x + 1$.

Similarly, we can find any irreducible polynomial of any degree in any $\mathbb{F}_p$. Generally there will be multiple irreducible polynomials of a certain degree.

Definition 2.2.7. We write $\mathbb{F}_{p^d}$ for a field with p^d elements. These fields are also known as *Galois fields* of order p^d.

A finite field with p^d elements will consist of all polynomials in $\mathbb{F}_p[x]$ with degree $< d$. Addition and multiplication will be done modulo p and modulo a *generating polynomial*. The generating polynomial will be an irreducible polynomial in $\mathbb{F}_p[x]$ with degree d.

Example 2.2.8. Let us construct the finite field $\mathbb{F}_{2^2}$ and the associated addition and multiplication tables. As seen in example 2.2.6, in $\mathbb{F}_{2^2}$ the polynomials of degree < 2 are 0, 1, x and $x+1$ and these constitute the elements of the finite field. The only irreducible polynomial of degree 2 is $x^2 + x + 1$ and it will act as the generating polynomial. When applying addition or multiplication the coefficients of the polynomials are calculated (mod 2) and the polynomials as a whole are calculated (mod $x^2 + x + 1$). In general the notation for equality will be used in these calculations. However, sometimes we will use notation for congruence in order to highlight some aspects of the calculations. First, let us look at addition in $\mathbb{F}_{2^2}$.

$$\begin{aligned} (x+1) + 1 &= x + 2 = x \\ x + (x+1) &= 2x + 1 = 1 \\ (x+1) + (x+1) &= 2x + 2 = 0 \end{aligned}$$

These equalities are true because $2 = 0$ in $\mathbb{F}_2$. The complete addition table is compiled in table 2.3. The central aspect of multiplication in a finite field is the

$+$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

Table 2.3: Addition table for $\mathbb{F}_{2^2}$.

generating polynomial. Therefore, we start with examining what information we can gain directly from the generating polynomial.

$$x^2 + x + 1 \equiv 0 \pmod{x^2 + x + 1}$$
$$x^2 \equiv -x - 1 \equiv x + 1 \pmod{x^2 + x + 1}$$

The congruence is true because $-1 \equiv 1 \pmod 2$. And therefore $-x \equiv x \pmod{x^2 + x + 1}$. With this in mind we can now perform the multiplications.

$$x(x+1) = x^2 + x = (x+1) + x = 2x + 1 = 1$$
$$(x+1)(x+1) = x^2 + 2x + 1 = (x+1) + 0 + 1 = x + 2 = x$$

The complete multiplication table is compiled in table 2.4.

$\cdot$	1	x	$x+1$
1	1	x	$x+1$
x	x	$x+1$	1
$x+1$	$x+1$	1	x

Table 2.4: Multiplication table for $\mathbb{F}_{2^2}$.

Example 2.2.9. Let us construct $\mathbb{F}_{2^3}$. $\mathbb{F}_2[x]$ has two irreducible polynomials of degree 3, x^3+x+1 and x^3+x^2+1. Thus, the multiplication tables for the finite field $\mathbb{F}_{2^3}$ will look different depending on which polynomial we choose, although they are isomorphic. The addition table will look the same regardless of the chosen generating polynomial. As for $\mathbb{F}_{2^2}$, the terms or coefficients are (mod 2) and the polynomials will be (mod *the generating polynomial*). The addition table is compiled in table 2.5.

$+$	0	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
0	0	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
1	1	0	$x+1$	x	x^2+1	x^2	x^2+x+1	x^2+x
x	x	$x+1$	0	1	x^2+x	x^2+x+1	x^2	x^2+1
$x+1$	$x+1$	x	1	0	x^2+x+1	x^2+x	x^2+1	x^2
x^2	x^2	x^2+1	x^2+x	x^2+x+1	0	1	x	$x+1$
x^2+1	x^2+1	x^2	x^2+x+1	x^2+x	1	0	$x+1$	x
x^2+x	x^2+x	x^2+x+1	x^2	x^2+1	x	$x+1$	0	1
x^2+x+1	x^2+x+1	x^2+x	x^2+1	x^2	$x+1$	x	1	0

Table 2.5: Addition table for $\mathbb{F}_{2^3}$.

With generating polynomial x^3+x+1 we have:

$$
\begin{aligned}
x^3 &= -x-1 = x+1 \qquad \text{because } 1=-1 \text{ in } \mathbb{F}_2\\
x^3+x^2 &= (x+1)+x^2 = x^2+x+1\\
x^3+x &= (x+1)+x = 1\\
x^3+1 &= (x+1)+1 = x\\
x^4 &= x\cdot x^3 = x(x+1) = x^2+x\\
x^4+x^3 &= (x^2+x)+(x+1) = x^2+2x+1 = x^2+1\\
x^4+x^2 &= (x^2+x)+x^2 = 2x^2+x = x\\
x^4+x &= (x^2+x)+x = x^2+2x = x^2\\
x^4+1 &= (x^2+x)+1 = x^2+x+1
\end{aligned}
$$

The result is compiled in table 2.6.

$\cdot$	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
1	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
x	x	x^2	x^2+x	$x+1$	1	x^2+x+1	x^2+1
$x+1$	$x+1$	x^2+x	x^2+1	x^2+x+1	x^2	1	x
x^2	x^2	$x+1$	x^2+x+1	x^2+x	x	x^2+1	1
x^2+1	x^2+1	1	x^2	x	x^2+x+1	$x+1$	x^2+x
x^2+x	x^2+x	x^2+x+1	1	x^2+1	$x+1$	x	x^2
x^2+x+1	x^2+x+1	x^2+1	x	1	x^2+x	x^2	$x+1$

Table 2.6: Multiplication table for $\mathbb{F}_{2^3}$ with generating polynomial x^3+x+1.

If we instead choose $x^3 + x^2 + 1$ as the generating polynomial we have:

$$\begin{aligned}
x^3 &= -x^2 - 1 = x^2 + 1 \qquad \text{because } 1 = -1 \text{ in } \mathbb{F}_2 \\
x^3 + x^2 &= (x^2 + 1) + x^2 = 1 \\
x^3 + x &= (x^2 + 1) + x = x^2 + x + 1 \\
x^3 + 1 &= (x^2 + 1) + 1 = x^2 + 2 = x^2 \\
x^4 &= x \cdot x^3 = x(x^2 + 1) = x^3 + x = (x^2 + 1) + x = x^2 + x + 1 \\
x^4 + x^3 &= (x^2 + x + 1) + (x^2 + 1) = 2x^2 + x + 2 = x \\
x^4 + x^2 &= (x^2 + x + 1) + x^2 = 2x^2 + x + 1 = x + 1 \\
x^4 + x &= (x^2 + x + 1) + x = x^2 + 2x + 1 = x^2 + 1 \\
x^4 + 1 &= (x^2 + x + 1) + 1 = x^2 + x
\end{aligned}$$

The result is compiled in table 2.7.

$\cdot$	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
1	1	x	$x+1$	x^2	x^2+1	x^2+x	x^2+x+1
x	x	x^2	x^2+x	x^2+1	x^2+x+1	1	$x+1$
$x+1$	$x+1$	x^2+x	x^2+1	1	x	x^2+x+1	x^2
x^2	x^2	x^2+1	1	x^2+x+1	$x+1$	x	x^2+x
x^2+1	x^2+1	x^2+x+1	x	$x+1$	x^2+x	x^2	1
x^2+x	x^2+x	1	x^2+x+1	x	x^2	$x+1$	x^2+1
x^2+x+1	x^2+x+1	$x+1$	x^2	x^2+x	1	x^2+1	x

Table 2.7: Multiplication table for $\mathbb{F}_{2^3}$ with generating polynomial $x^3 + x^2 + 1$.

Example 2.2.10. There are three irreducible polynomials of degree 2 in $\mathbb{F}_3[x]$. As in the previous example, the addition table for the finite field $\mathbb{F}_{3^2}$ will be the same regardless of the generating polynomial chosen, see table 2.8, but the multiplication tables will be different. Addition is (mod 3) and multiplication is modulo the chosen generating polynomial.

With generating polynomial $x^2 + 1$ we have that

$$x^2 \equiv -1 \equiv 2 \pmod{x^2 + 1} \qquad \text{because } -1 = 2 \text{ in } \mathbb{F}_3.$$

$+$	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
0	0	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
1	1	2	0	$x+1$	$x+2$	x	$2x+1$	$2x+2$	$2x$
2	2	0	1	$x+2$	x	$x+1$	$2x+2$	$2x$	$2x+1$
x	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$	0	1	2
$x+1$	$x+1$	$x+2$	x	$2x+1$	$2x+2$	$2x$	1	2	0
$x+2$	$x+2$	x	$x+1$	$2x+2$	$2x$	$2x+1$	2	0	1
$2x$	$2x$	$2x+1$	$2x+2$	0	1	2	x	$x+1$	$x+2$
$2x+1$	$2x+1$	$2x+2$	$2x$	1	2	0	$x+1$	$x+2$	x
$2x+2$	$2x+2$	$2x$	$2x+1$	2	0	1	$x+2$	x	$x+1$

Table 2.8: Addition table for $\mathbb{F}_{3^2}$.

And therefore

$$\begin{aligned} x^2 + x &= x - 1 = x + 2 & 2x^2 &= x^2 - 1 = x^2 + 2 = 1 \\ x^2 + 2x &= 2x - 1 = 2x + 2 & 2x^2 + x &= x + 1 \\ x^2 + 1 &= 0 & 2x^2 + 2x &= 2x + 1 \\ x^2 + 2 &= 1 & 2x^2 + 1 &= 2 \\ & & 2x^2 + 2 &= 0. \end{aligned}$$

The result is compiled in table 2.9.

$\cdot$	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
1	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
2	2	1	$2x$	$2x+2$	$2x+1$	x	$x+2$	$x+1$
x	x	$2x$	2	$x+2$	$2x+2$	1	$x+1$	$2x+1$
$x+1$	$x+1$	$2x+2$	$x+2$	$2x$	1	$2x+1$	2	x
$x+2$	$x+2$	$2x+1$	$2x+2$	1	x	$x+1$	$2x$	2
$2x$	$2x$	x	1	$2x+1$	$x+1$	2	$2x+2$	$x+2$
$2x+1$	$2x+1$	$x+2$	$x+1$	2	$2x$	$2x+2$	x	1
$2x+2$	$2x+2$	$x+1$	$2x+1$	x	2	$x+2$	1	$2x$

Table 2.9: Multiplication table for $\mathbb{F}_{3^2}$ with generating polynomial x^2+1.

With generating polynomial $x^2 + x + 2$ we have that

$$x^2 \equiv -x - 2 \equiv 2x + 1 \pmod{x^2 + x + 2}$$
$$\text{because } -1 = 2 \text{ and } -2 = 1 \text{ in } \mathbb{F}_3.$$

And therefore

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

The result is compiled in table 2.10.

$\cdot$	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
1	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
2	2	1	$2x$	$2x+2$	$2x+1$	x	$x+2$	$x+1$
x	x	$2x$	$2x+1$	1	$x+1$	$x+2$	$2x+2$	2
$x+1$	$x+1$	$2x+2$	1	$x+2$	$2x$	2	x	$2x+1$
$x+2$	$x+2$	$2x+1$	$x+1$	$2x$	2	$2x+2$	1	x
$2x$	$2x$	x	$x+2$	2	$2x+2$	$2x+1$	$x+1$	1
$2x+1$	$2x+1$	$x+2$	$2x+2$	x	1	$x+1$	2	$2x$
$2x+2$	$2x+2$	$x+1$	2	$2x+1$	x	1	$2x$	$x+2$

Table 2.10: Multiplication table for $\mathbb{F}_{3^2}$ with generating polynomial $x^2 + x + 2$.

With generating polynomial $x^2 + 2x + 2$ we have that

$$x^2 \equiv -2x - 2 \equiv x + 1 \pmod{x^2 + 2x + 2} \qquad \text{because } -2 = 1 \text{ in } \mathbb{F}_3.$$

And therefore

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

The result is compiled in table 2.11.

$\cdot$	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
1	1	2	x	$x+1$	$x+2$	$2x$	$2x+1$	$2x+2$
2	2	1	$2x$	$2x+2$	$2x+1$	x	$x+2$	$x+1$
x	x	$2x$	$x+1$	$2x+1$	1	$2x+2$	2	$x+2$
$x+1$	$x+1$	$2x+2$	$2x+1$	2	x	$x+2$	$2x$	1
$x+2$	$x+2$	$2x+1$	1	x	$2x+2$	2	$x+1$	$2x$
$2x$	$2x$	x	$2x+2$	$x+2$	2	$x+1$	1	$2x+1$
$2x+1$	$2x+1$	$x+2$	2	$2x$	$x+1$	1	$2x+2$	x
$2x+2$	$2x+2$	$x+1$	$x+2$	1	$2x$	$2x+1$	x	2

Table 2.11: Multiplication table for $\mathbb{F}_{3^2}$ with generating polynomial x^2+2x+2.

As mentioned in 2.2.5 all finite fields with a given number of elements are isomorphic. To illustrate this, let us look at the previous examples of $\mathbb{F}_{2^3}$ and $\mathbb{F}_{3^2}$.

Example 2.2.11. Let X be the finite field $\mathbb{F}_{2^3}$ with generating polynomial x^3+x+1 and let Y be the finite field $\mathbb{F}_{2^3}$ with generating polynomial y^3+y^2+1. See addition and multiplication tables 2.5, 2.6 and 2.7. To find how X maps to Y one must find a relation between one element of X and one element of Y. Let $x = ay^2 + by + c$ where $a, b, c \in \mathbb{F}_2 = \{0, 1\}$.

$$\begin{aligned} x^3 &= x + 1 \\ (ay^2 + by + c)^3 &= (ay^2 + by + c) + 1 \end{aligned}$$

Expanding the left hand side of the equation yields:

$$(ay^2+by+c)^3 = a^3y^6+3a^2by^5+(3a^2c+3ab^2)y^4+(6abc+b^3)y^3+ \\ +(3ac^2+3b^2c)y^2+3bc^2y+c^3.$$

Since we are in $\mathbb{F}_{2^3}$ all even coefficients are equal to 0 and all odd coefficients are equal to 1. We also know from previous examples that $y^3 = y^2+1$ and $y^4 = y^2+y+1$. From that we deduce that $y^5 = y+1$ and that $y^6 = y^2+y$. With this information the equation becomes:

$$\ldots = (a^3+a^2c+ab^2+b^3+ac^2+b^2c)y^2+ \\ +(a^3+a^2b+a^2c+ab^2+bc^2)y+ \\ +a^2b+a^2c+ab^2+b^3+c^3$$

Comparing the left hand side of the original equation with the right hand side we see that

$$a^3+a^2c+ab^2+b^3+ac^2+b^2c = a \\ a^3+a^2b+a^2c+ab^2+bc^2 = b \\ a^2b+a^2c+ab^2+b^3+c^3 = c+1.$$

We can solve these equations to acquire the values for a, b and c but since $a,b,c = 0$ or 1 there are only 8 different combination and it may be easier to try which of those solves all three equations. There are three distinct solutions, $(a,b,c) = (1,1,0)$, $(a,b,c) = (1,0,1)$ and $(a,b,c) = (0,1,1)$. So,

$$x_1 = y^2+y \\ x_2 = y^2+1 \\ x_3 = y+1$$

are all mappings of X into Y.
The remainder of this example will be limited to only one of these solutions as the process to find the remaining mappings of elements will be similar regardless of the chosen solution.
For $x_1 = y^2+y$ we get

$$x_1+1 = y^2+y+1 \\ x_1^2 = (y^2+y)^2 = y^4+2y^3+y^2 = y+1 \\ x_1^2+1 = (y+1)+1 = y \\ x_1^2+x_1 = (y+1)+(y^2+y) = y^2$$

and 0 and 1 maps onto themselves.
To confirm that this is indeed an isomorphic mapping from X to Y one can insert the elements of Y into the addition and multiplications tables of X according to the mappings above and then rearrange the rows and columns. It is not difficult and will be left to the reader.

Example 2.2.12. Let X be the finite field $\mathbb{F}_{3^2}$ with generating polynomial x^2+1, Y the finite field $\mathbb{F}_{3^2}$ with generating polynomial y^2+y+2 and Z the finite field $\mathbb{F}_{3^2}$ with generating polynomial z^2+2z+2. See the addition and multiplication tables 2.8, 2.9, 2.10 and 2.11.
First, let us find the mapping from X to Y. Let $x = ay + b$ where $a, b \in \mathbb{F}_3 = \{0, 1, 2\}$.

$$x^2 = 2$$
$$(ay+b)^2 = 2$$

Expanding the left hand side of the equation, and taking into account that $y^2 = 2y + 1$ and that the coefficients are (mod 3), yield

$$(ay+b)^2 = a^2y^2 + 2aby + b^2 = a^2(2y+1) + 2aby + b^2 = \\ = (2a^2 + 2ab)y + a^2 + b^2 = 2.$$

Comparing the left hand side and the right hand side yields:

$$2a^2 + 2ab = 0$$
$$a^2 + b^2 = 2.$$

Here there are 9 different possible combinations for a and b. The two solutions are $(a, b) = (2, 1)$ and $(a, b) = (1, 2)$. Therefore,

$$x_{y1} = 2y + 1$$
$$x_{y2} = y + 2.$$

A complete mapping of X and Y is calculated from any of these relations.
Next, let us find the mapping from X to Z. Let $x = cz + d$ where $c, d \in \mathbb{F}_3 = \{0, 1, 2\}$.

$$x^2 = 2$$
$$(cz+d)^2 = 2$$

Expanding the left hand side of the equation, and taking into account that $z^2 = z + 1$ and that the coefficients are (mod 3), yield

$$(cz+d)^2 = c^2y^2 + 2cdy + d^2 = c^2(z+1) + 2cdz + d^2 = \\ = (c^2 + 2cd)z + c^2 + d^2 = 2.$$

Comparing the left hand side and the right hand side yield:

$$c^2 + 2cd = 0$$
$$c^2 + d^2 = 2.$$

The two solutions are $(c, d) = (1, 1)$ and $(c, d) = (2, 2)$ so that

$$x_{z1} = z + 1$$
$$x_{z2} = 2z + 2.$$

To find the mapping between Y and Z we can follow the same process as above. But since we already have mappings from X to Y and Z, and all fields $\mathbb{F}_{3^2}$ are isomorphic, we can calculate the mappings from Y to Z from the ones we already have.

$$x_{y1} = 2y + 1 = z + 1 = x_{z1}$$
$$y = 2z$$
and
$$x_{y1} = 2y + 1 = 2z + 2 = x_{z2}$$
$$y = z + 2.$$

So, the mappings from Y to Z are

$$y_{z1} = 2z$$
$$y_{z2} = z + 2.$$

2.3 Resultants and Discriminants

Resultants and discriminants are determinants containing coefficients from polynomials and they are commonly used to determine if two polynomials have any common factors. Our main interest is that the relations presented here are used later to support and prove concepts related to intersections of curves. The source used in this section is [6].

Theorem 2.3.1. *f and g have a common non-constant factor if and only if there exist non-zero polynomials ϕ and ψ, where $\deg \phi < \deg f$ and $\deg \psi < \deg g$, such that $\psi f = \phi g$.*

Example 2.3.2. Let $f(x) = (x - 1)(x + 1)(x + 2)$ and $g(x) = (x - 1)(x - 2)$. Then $\phi(x) = (x + 1)(x + 2)$ and $\deg \phi = 2 < 3 = \deg f$ and $\psi(x) = (x - 2)$ and $\deg \psi = 1 < 2 = \deg g$.

If f and g lack a common factor, ϕ and ψ must be equal to g and f respectively for the equality $\psi f = \phi g$ to hold, which is a contradiction to the degree criterion.

Theorem 2.3.3. *Let* $f = a_0x^n + a_1x^{n-1} + \ldots + a_nx^0$ *and* $g = b_0x^m + b_1x^{m-1} + \ldots + b_mx^0$, *where* $a_0 \neq 0$ *and* $b_0 \neq 0$. f *and* g *have a common non-constant factor if and only if*

$$\mathrm{Res}(f,g,x) = \begin{vmatrix} a_0 & a_1 & \ldots & \ldots & a_n & & & \\ & a_0 & \ldots & \ldots & a_{n-1} & a_n & & \\ & & \ddots & & & & \ddots & \\ & & & a_0 & \ldots & \ldots & \ldots & a_n \\ b_0 & b_1 & \ldots & b_{m-1} & b_m & & & \\ & & \ddots & & & \ddots & & \\ & & & b_0 & \ldots & \ldots & \ldots & b_m \end{vmatrix} = 0$$

there being m *rows of* a*'s and* n *rows of* b*'s in the determinant, the rows being filled out with zeros.* $\mathrm{Res}(f,g,x)$ *is called the resultant of* f *and* g *(with respect to* x*). The resultant of a polynomial and its derivative,* f'*, is called the discriminant of the polynomial, denoted* $\mathrm{Disc}(f,x)$.

Remark. In different literature the resultant is sometimes defined as the transpose of the determinant stated above. Either is fine since $\det(A) = \det(A^t)$ and it is merely a case of different practices.

Example 2.3.4. Let $f = a_0x^3 + a_1x^2 + a_2x + a_3$ and $g = b_0x^2 + b_1x + b_2$, where $a_0 \neq 0$ and $b_0 \neq 0$. So $\deg(f) = 3$ and $\deg(g) = 2$. This means that there will be two rows of f:s coefficients and three rows of g:s coefficients in their resultant determinant. So

$$\mathrm{Res}(f,g,x) = \begin{vmatrix} a_0 & a_1 & a_2 & a_3 & 0 \\ 0 & a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & 0 & 0 \\ 0 & b_0 & b_1 & b_2 & 0 \\ 0 & 0 & b_0 & b_1 & b_2 \end{vmatrix}.$$

Example 2.3.5. Let $f = x^2 + x - 2 = (x-1)(x+2)$ and $g = x - 1$. Then the

resultant

$$\begin{aligned}\operatorname{Res}(f,g,x) &= \begin{vmatrix} 1 & 1 & -2 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{vmatrix} = \\ &= (1\cdot(-1)\cdot(-1)) - (1\cdot 0\cdot 1) + (1\cdot 0\cdot 0) - (1\cdot 1\cdot -1) + \\ &\qquad + (-2\cdot 1\cdot 1) - (-2\cdot -1\cdot 0) = \\ &= 1+1-2=0.\end{aligned}$$

Example 2.3.6. Let $f = x^3 + 2x^2 - x - 2 = (x-1)(x+1)(x+2)$ and $g = x^2 - 3x + 2 = (x-1)(x-2)$. Then the resultant

$$\operatorname{Res}(f,g,x) = \begin{vmatrix} 1 & 2 & -1 & -2 & 0 \\ 0 & 1 & 2 & -1 & -2 \\ 1 & -3 & 2 & 0 & 0 \\ 0 & 1 & -3 & 2 & 0 \\ 0 & 0 & 1 & -3 & 2 \end{vmatrix} = \ldots = 2\cdot 12^2 - 3\cdot 12^2 + 12^2 = 0.$$

Example 2.3.7. Let $f = x^2 + x - 2 = (x-1)(x+2)$ and $g = x - 2$. Then the resultant

$$\operatorname{Res}(f,g,x) = \begin{vmatrix} 1 & 1 & -2 \\ 1 & -2 & 0 \\ 0 & 1 & -2 \end{vmatrix} = 4 + 2 - 2 = 4 \neq 0.$$

Theorem 2.3.8. *A polynomial over a domain of characteristic zero has a repeated non-constant factor if and only if its discriminant is zero.*

Example 2.3.9. Let $f = x^3 - x^2 - x + 1 = (x-1)^2(x+1)$ and its derivative $f' = 3x^2 - 2x - 1 = (3x+1)(x-1)$. Then the discriminant

$$\operatorname{Disc}(f,x) = \begin{vmatrix} 1 & -1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 & 1 \\ 3 & -2 & -1 & 0 & 0 \\ 0 & 3 & -2 & -1 & 0 \\ 0 & 0 & 3 & -2 & -1 \end{vmatrix} = \ldots = 0.$$

Example 2.3.10. Let $f = x^3 + 2x^2 - x - 2 = (x-1)(x+1)(x+2)$ and its derivative $f' = 3x^2 + 4x - 1 = 3\left(x + \frac{2}{3} + \frac{\sqrt{7}}{3}\right)\left(x + \frac{2}{3} - \frac{\sqrt{7}}{3}\right)$. Then the

discriminant

$$\mathrm{Disc}(f,g,x) = \begin{vmatrix} 1 & 2 & -1 & -2 & 0 \\ 0 & 1 & 2 & -1 & -2 \\ 3 & 4 & -1 & 0 & 0 \\ 0 & 3 & 4 & -1 & 0 \\ 0 & 0 & 3 & 4 & -1 \end{vmatrix} = \ldots = 14 \cdot (-2) \cdot 7 - 10 \cdot (-2) \cdot 8 =$$

$$= -196 + 160 = -36 \neq 0.$$

It is also possible to determine existence of common factors in polynomials of several variables. However, we can only determine the existence of a common factor with respect to one variable at the time. Example 2.3.13 illustrate this.

Definition 2.3.11. Let $f, g \in K[x_1, \ldots, x_n]$, where

$$f = a_0 x_1^l + \ldots + a_l, \qquad a_0 \neq 0$$
$$g = b_0 x_1^m + \ldots + b_m, \qquad b_0 \neq 0,$$

and $a_i, b_i \in K[x_2, \ldots, x_n]$. We define the *resultant* of f and g with respect to x_1 to be the determinant

$$\mathrm{Res}(f,g,x_1) = \begin{vmatrix} a_0 & a_1 & \ldots & \ldots & a_l & & & \\ & a_0 & \ldots & \ldots & a_{l-1} & a_l & & \\ & & \ddots & & & & \ddots & \\ & & & a_0 & \ldots & \ldots & \ldots & a_l \\ b_0 & b_1 & \ldots & b_{m-1} & b_m & & & \\ & & \ddots & & & \ddots & & \\ & & & b_0 & \ldots & \ldots & \ldots & b_m \end{vmatrix}$$

where there are m rows of a_i and l rows of b_i and where empty spaces are filled with zeros.

Proposition 2.3.12. Let $f, g \in K[x_1, \ldots, x_n]$ have positive degrees in x_1. Then $\mathrm{Res}(f,g,x_1) = 0$ if and only if f and g have a common factor in $K[x_1, \ldots, x_n]$ which has a positive degree in x_1.

Example 2.3.13. Let $f(x,y) = x^2y + x^2 + xy^2 - x - y - y^2 = (x+y)(x-1)(y+1)$ and $g(x,y) = xy - x - y + 1 = (x-1)(y-1)$. If we view x as the only variable and y as a constant, we can write f and g as $f = (y+1)x^2 + (y^2-1)x^1 + (-y^2-y)x^0$

and $g = (y-1)x^1 + (-y+1)x^0$. The resultant of f and g with respect to x is

$$\begin{aligned}\mathrm{Res}(f,g,x) &= \begin{vmatrix} y+1 & y^2-1 & -y^2-y \\ y-1 & -y+1 & 0 \\ 0 & y-1 & -y+1 \end{vmatrix} = \\ &= (y+1)(1-y)^2 - (y^2-1)(y-1)(1-y) + (-y^2-y)(y-1)^2 = \\ &= (y^3-y^2-y+1) + (y^4-2y^3+2y-1) + (-y^4+y^3+y^2-y) = \\ &= 0.\end{aligned}$$

If we instead view y as the only variable and x as a constant, we can write f and g as $f = (x-1)y^2 + (x^2-1)y^1 + (x^2+x)y^0$ and $g = (x-1)y^1 + (-x+1)y^0$. The resultant of f and g with respect to y is

$$\begin{aligned}\mathrm{Res}(f,g,y) &= \begin{vmatrix} x-1 & x^2-1 & x^2-x \\ x-1 & -x+1 & 0 \\ 0 & x-1 & -x+1 \end{vmatrix} = \\ &= (x-1)(1-x)^2 - (x^2-1)(x-1)(1-x) + (x^2-x)(x-1)^2 = \\ &= (x^3-3x^2+3x-1) + (x^4-2x^3+2x-1) + (x^4-3x^3+3x^2-x) = \\ &= 2x^4-4x^3+4x-2 \neq 0.\end{aligned}$$

We see that a resultant that is non-zero does not necessarily mean that there is no common factor for the polynomials, just that there is no common factor containing the chosen variable.

2.4 Projective Spaces

The reader is probably familiar with some concept of infinity, for example when working with limits in calculus. Usually, one can not assign a fixed value or coordinates to the infinity. However, when working with projective spaces infinity gets assigned coordinates. This will be essential later when discussing curves and intersections between curves in section 2.6. The sources used in this section are [1] and [6].

Definition 2.4.1. Let $\sim$ be an equivalence class of non-zero points on the field K^{n+1} such that

$$(x_0', \ldots, x_n') \sim (x_0, \ldots, x_n)$$

if there $\exists \lambda \in K \setminus \{0\}$ such that $(x_0', \ldots, x_n') = \lambda(x_0, \ldots, x_n)$. The *$n$-dimensional projective space over the field K*, denoted $\mathbb{P}^n(K)$, is the set of equivalence classes of $\sim$ on $K^{n+1} - \{0\}$. Thus

$$\mathbb{P}^n(K) = (K^{n+1} - \{0\}) / \sim .$$

Each non-zero $(n+1)$-tuple $(x_0, \ldots, x_n) \in K^{n+1}$ defines a point P in $\mathbb{P}^n(K)$, and we say that $(x_0 : \ldots : x_n)$ are *homogeneous coordinates* of P.

Remark. In some literature the name projective coordinates are used instead.

To understand what this definition means in practice we shall look at a few different examples that have slightly different approaches.

Example 2.4.2. Every point, P, on an arbitrary line can be described with a single coordinate $P = (a)$, where $a \in \mathbb{R}$, except for the point at infinity. To remedy this, imagine letting lines trough the origin intersect the line $y = 1$ in the xy-plane. All points on $y = 1$ can now be represented by the directional vector, of their respective intersecting line. All points, except the point at infinity, can be represented by coordinates of the form $(a:1)$, (because $(s:t) \sim \left(\frac{s}{t}:1\right)$). If we let the intersection of the lines approach infinity we will see that the line through the origin will approach the x-axis. Therefore we define the coordinates for the point at infinity to be $(1:0)$. In this way we have constructed the projective line $\mathbb{P}(\mathbb{R})$.

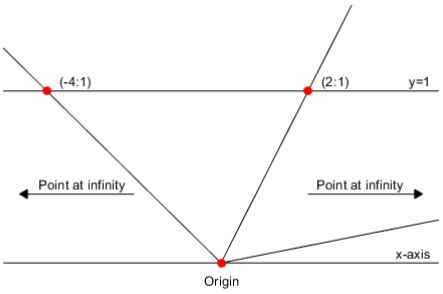

Figure 2.1: Construction of the projective line $\mathbb{P}(\mathbb{R})$.

It is easy to think that there is only one point at infinity, and in the one dimensional case that is true. But in dimensions greater than one there will be infinitely many points at infinity. In two dimensions the points at infinity are said to lie on *the line at infinity*.

Example 2.4.3. The projective plane, $\mathbb{P}^2(\mathbb{R})$, can be constructed in a similar way as the projective line, $\mathbb{P}(\mathbb{R})$. Let lines through the origin intersect the plane

$z = 1$. All points on $z = 1$, except the points at infinity, can be represented by the intersecting lines' directional vectors. This will yield coordinates of the form $(a : b : 1)$, where $a, b \in \mathbb{R}$. Lines that approaches infinity in any direction will approach their respective line in the xy-plane. Therefore, any point at infinity will have coordinates of the form $(a : b : 0)$. Another way to understand the

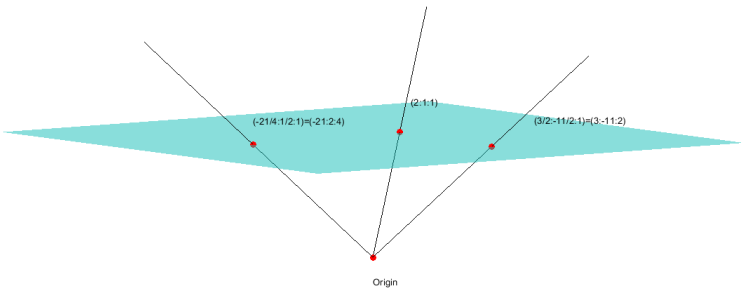

Figure 2.2: Construction of the projective plane $\mathbb{P}^2(\mathbb{R})$.

projective plane, $\mathbb{P}^2(\mathbb{R})$, is the real plane, $\mathbb{R}^2$, with one extra point at infinity for each equivalence class of parallel lines. This is equivalent to the previous construction because parallel lines have the same directional vectors regardless of where in the plane they lie. Therefore they will have the coordinates stated above. The projective lines in $\mathbb{P}^2(\mathbb{R})$ are described by equations of the form

$$Ax + By + Cz = 0, \qquad A, B, C \text{ not all zero.}$$

In summary, points and lines in $\mathbb{P}^2(\mathbb{R})$ can be viewed as lines respectively planes through the origin in $\mathbb{R}^3$.

Example 2.4.4. Another way to construct the projective line is to use stereographic projection. Let $S : x^2 + y^2 - y = 0$ be a circle in the xy-plane and let lines extend from its "north pole", $(0, 1)$, and intersect the x-axis. Every point on the x-axis can now be identified with a point on the circle, specifically the point at infinity can be represented by the "north pole", $(0 : 1)$.

The example should feel familiar to anyone who has encountered the Riemann sphere, because they are the same kind of relation.

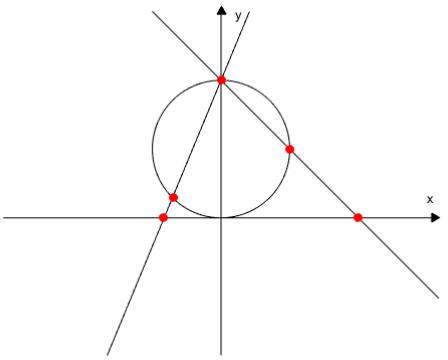

Figure 2.3: Stereographic projection of the projective line.

Example 2.4.5. The relation between the complex plane and the Riemann sphere is an example of a complex projective line, $\mathbb{P}(\mathbb{C})$, and not a projective plane, despite the name "complex plane". The Riemann sphere is a unit sphere. Points in the complex plane where $|z| = 1$ is represented by points on the sphere's equator, points where $|z| < 1$ is represented on the sphere's lower half and points where $|z| > 1$ is represented on the sphere's upper half. Specifically, the point at infinity is represented by the sphere's "north pole", $(0:0:1)$.

Definition 2.4.6. Given a coordinate system S, a coordinate system S' is determined by associating any point P, with coordinates $(x) = (x_0 : x_1 : \ldots : x_n)$ in S, with the coordinates $(y) = (y_0 : y_1 : \ldots : y_n)$ in S', where

$$y_i = \sum_{j=0}^{n} \alpha_{i,j} x_j = \alpha_{i,0} x_0 + \alpha_{i,1} x_1 + \ldots + \alpha_{i,n} x_n,$$

and $\alpha_{i,j}$, $i, j = 0, \ldots, n$ are numbers such that $|\alpha_{i,j}| = \alpha \neq 0$.

What this means in practice is that it is possible to change coordinate system to one suitable to one's current calculations or when conducting proofs.

2.5 Algebraic Curves

As hinted in the previous section, projective spaces are used when analysing intersections of curves. This is because some intersections can occur at infinity

and in order to get the number of intersections correct these must be taken into account. Before we can do that we have to define what a curve is and the differences and similarities between its homogeneous and affine representations. To do this, we will begin by introducing homogeneous polynomials and some of their properties. The source used in this section is [6].

Definition 2.5.1. A polynomial f is said to be *homogeneous* if all of its terms are of the same degree.

Example 2.5.2. The polynomial $f(x, y) = 4x^2y^3 + xy^4 - 2y^5$ is homogeneous of degree 5. The polynomial $g(x, y) = x^5 - 4x^2y + y^4$ is non-homogeneous but also of degree 5.

Theorem 2.5.3. *$f(x_1, \ldots, x_r) \neq 0$ is homogeneous of degree n if and only if the relation*

$$f(tx_1, \ldots, tx_r) = t^n f(x_1, \ldots, x_r)$$

holds between these elements of $D[x_1, \ldots, x_r]$.

It is common to denote homogeneous polynomials with upper-case letters. F_i generally denotes a homogeneous polynomial of degree i. These practices will be followed throughout this essay. The proof of theorem 2.5.3 is shown with an easy calculation from [6].

Proof. Let

$$f = F_{n_1} + F_{n_2} + \ldots + F_{n_k}, \qquad n_1 < n_2 < \ldots < n_k$$

where F_{n_i} is homogeneous of degree n_i and is not zero. Then $f(tx_1, \ldots, tx_r) = t^n f(x_1, \ldots, x_r)$ becomes

$$t^{n_1}F_{n_1} + t^{n_2}F_{n_2} + \ldots + t^{n_k}F_{n_k} = t^n F_{n_1} + t^n F_{n_2} + \ldots + t^n F_{n_k}.$$

Hence we must have $t^{n_i} = t^n$ for each i. This implies that $k = 1$ and $n_1 = n$. □

Definition 2.5.4. Let $F_n \in D[x_0, \ldots, x_r]$ be a homogeneous polynomial of degree n. If $x_0 \nmid F_n$ the polynomial $f(x_1, \ldots, x_r) = F(1, x_1, \ldots, x_r)$ is also of degree n. f and F_n are said to be *associated*. Conversely, given any polynomial $f(x_1, \ldots, x_r)$ of degree n we can obtain its associated homogeneous polynomial by introducing the proper powers of x_0 into those terms of f which are of degree less than n.

Example 2.5.5. Let $F(x,y,z) = 2x^5 + y^2z^3 - 5xy^3z + xz^4$. Then

$$\begin{aligned} F(tx,ty,tz) &= 2(tx)^5 + (ty)^2(tz)^3 - 5(tx)(ty)^3(tz) + (tx)(tz)^4 = \\ &= 2t^5x^5 + t^5y^2z^3 - 5t^5xy^3z + t^5xz^4 = \\ &= t^5F(x,y,z). \end{aligned}$$

Neither x, y or z divides F, therefore there are three different associates to F. The associates of F are

$$\begin{aligned} f_1(x,y) &= F(x,y,1) = 2x^5 + y^2 - 5xy^3 + x \\ f_2(x,z) &= F(x,1,z) = 2x^5 + z^3 - 5xz + xz^4 \\ f_3(y,z) &= F(1,y,z) = 2 + y^2z^3 - 5y^3z + z^4. \end{aligned}$$

Theorem 2.5.6. *If F_n and f are associated, any factor of F_n is associated with a factor of f and vice versa.*

Theorem 2.5.7. *Any factor of a homogeneous polynomial is homogeneous.*

Again, we look at the proof from [6].

Proof. Let $F = fg$ and suppose that f is not homogeneous. Then we may write

$$f = F_i + F_{i+1} + \ldots + F_{i+j},$$

where $F_i \neq 0$, $F_{i+j} \neq 0$, and $j > 0$. Similarly,

$$g = G_k + G_{k+1} + \ldots + G_{k+l},$$

where $G_k \neq 0$, $G_{k+l} \neq 0$ and $l \geq 0$. Then

$$F = fg = F_iG_k + (F_{i+1}G_k + F_iG_{k+1}) + \ldots + F_{i+j}G_{k+l}.$$

Now $F_iG_k \neq 0$ and $F_{i+j}G_{k+l} \neq 0$ and

$$\deg F_iG_k = i + k < i + j + k + l = \deg F_{i+j}G_{k+l}.$$

Hence F is not homogeneous, contrary to assumption. □

Example 2.5.8. Let $F(x,y,z) = xy + xz - 2yz - 2z^2 = (x - 2z)(y + z)$. F is homogeneous and of degree 2 and F:s factors are homogeneous and of degree 1. F:s associates are

$$\begin{aligned} f_1(x,y) &= F(x,y,1) = xy + x - 2y - 2 = (x-2)(y+1) \\ f_2(x,z) &= F(x,1,z) = x + xz - 2z - 2z^2 = (x-2z)(z+1) \\ f_3(y,z) &= F(1,y,z) = y + z - 2yz - 2z^2 = (1-2z)(y+z). \end{aligned}$$

$x-2z$ associates are $x-2$, $x-2z$ and $1-2z$. $y+z$ associates are $y+1$, $z+1$ and $y+x$.

Theorem 2.5.9. *F is irreducible if and only if its associate is irreducible.*

Resultants for homogeneous polynomials function much the same as resultants for any other polynomial of one or more variables. The theorems and example will be similar to those presented in section 2.3.

Theorem 2.5.10. *The homogeneous polynomials*

$$F = a_0x_0^n + a_1x_0^{n-1}x_1 + \ldots + a_nx_1^n,$$
$$G = b_0x_0^m + b_1x_0^{m-1}x_1 + \ldots + b_mx_1^m,$$

have a common non-constant factor with respect to x_i, $i = 0, 1$, if and only if their resultant, $\mathrm{Res}(F, G, x_i)$*, is zero.*

Theorem 2.5.11. *Let*

$$F_n = A_n + A_{n-1}x_r + \ldots + A_0x_r^n,$$
$$G_m = B_m + B_{m-1}x_r + \ldots + B_0x_r^m,$$

where A_i, B_i are homogeneous polynomials of degree i in $x_1, \ldots, x_{r-1}$, and $A_0B_0 \neq 0$. The resultant of F and G with respect to x_r, $\mathrm{Res}(F, G, x_r)$*, is either equal to 0 or is a homogeneous polynomial of degree mn.*

Example 2.5.12. Let $F(x,y) = 3x^3-2x^2y-7xy^2-2y^3 = (x+y)(x-2y)(3x+y)$ and $G(x,y) = x^2 - 3xy - 4y^2 = (x+y)(x-4y)$. Then the resultant of F and G with respect to x is

$$\mathrm{Res}(F,G,x) = \begin{vmatrix} 3 & -2y & -7y^2 & -2y^3 & 0 \\ 0 & 3 & -2y & -7y^2 & -2y^3 \\ 1 & -3y & -4y^2 & 0 & 0 \\ 0 & 1 & -3y & -4y^2 & 0 \\ 0 & 0 & 1 & -3y & -4y^2 \end{vmatrix} = \ldots = 0.$$

Example 2.5.13. Let $F(x,y) = 3x^3-2x^2y-7xy^2-2y^3 = (x+y)(x-2y)(3x+y)$ and $H(x,y) = x^2 - 2xy - 8y^2 = (x+2y)(x-4y)$. Then the resultant of F and H with respect to x is

$$\mathrm{Res}(F,H,x) = \begin{vmatrix} 3 & -2y & -7y^2 & -2y^3 & 0 \\ 0 & 3 & -2y & -7y^2 & -2y^3 \\ 1 & -2y & -8y^2 & 0 & 0 \\ 0 & 1 & -2y & -8y^2 & 0 \\ 0 & 0 & 1 & -2y & -8y^2 \end{vmatrix} = \ldots = -2600y^6 \neq 0.$$

Now that we have established the concept of homogeneous polynomials and its associates, we are ready to move on to curves. We will find a similar correspondences between the *algebraic curve* and its associate *affine curve*.

Definition 2.5.14. Let $F(x) = F(x_0, x_1, \ldots, x_r) \in \mathbb{P}^r(K)$ be an irreducible homogeneous polynomial of degree $n > 0$. The set of all points $P = (a_0, a_1, \ldots, a_r)$ that satisfy the equation $F(x) = 0$ is called an *irreducible algebraic curve*, and the equation $F(x) = 0$ is the *equation of the curve* in the given coordinate system.

Observe that if one set of coordinates $(a) = (a_0, a_1, \ldots, a_r)$ for a point P satisfy $F(x) = 0$ then all sets will, since

$$F(\lambda a_0, \lambda a_1, \ldots, \lambda a_r) = \lambda^n F(a_0, a_1, \ldots, a_r).$$

Definition 2.5.15. Let $F(x) = F(x_0, x_1, \ldots, x_r)$ be a homogeneous polynomial of degree $n > 0$. $F(x)$ has a unique factorisation, $F = F_1 F_2 \cdot F_s$, where $F_1, F_2, \ldots F_s$ are irreducible homogeneous polynomials that are unique up to constant multiples. The set of irreducible curves $C_1, \ldots, C_s$ whose equations are $F_1(x) = 0, \ldots, F_s(x) = 0$ is said to be an algebraic curve C whose equation is $F(x) = 0$, and the irreducible curves $C_1, \ldots, C_s$ are called *components* of C. An irreducible curve appearing more than once as a component of C is said to be a *multiple* component of C.

Thus, an irreducible curve is a curve with only one component and a curve with multiple components is said to be *reducible*.

Example 2.5.16. Let

$$\begin{aligned}
C_1 &: \quad F_1(x, y, z) = x - y + z = 0 \\
C_2 &: \quad F_2(x, y, z) = x^2 + yz = 0 \\
C_3 &: \quad F_3(x, y, z) = x^3 + y^2 z + yz^2 = 0 \\
C_4 &: \quad F_4(x, y, z) = (x - y + z)^2 = 0 \\
C_5 &: \quad F_5(x, y, z) = (x - y + z)(x + 2y - z) = 0.
\end{aligned}$$

Then C_1, C_2 and C_3 are irreducible curves of order 1, 2 and 3 respectively. C_4 is a reducible curve with a multiple component and C_5 is a reducible curve without multiple components.

As previously mentioned, in the same way that homogeneous polynomials have associates, projective algebraic curves have an affine algebraic curve counterpart.

Definition 2.5.17. Let $F(x) = 0$ be a curve C where $x_0 \nmid F$, that is, x_0 is not a component of C. Then there is an associated non-homogeneous polynomial $f(x_1, \ldots, x_r) = F(1, x_1, \ldots, x_r)$ where $f(x_1, \ldots, x_r) = 0$ is the equation of the curve C in the corresponding affine coordinate system. It is said to be the *affine part of the curve.*

2.6 Intersections of Curves

In this section we will examine the intersections of two curves, first the intersections of a curve and a line and then two curves of arbitrary order. We will alternate between a projective space $\mathbb{P}^2(K)$, where K is a closed fields of characteristic zero, and its affine counterpart. Polynomials denoted in upper-case letter, $F(x)$, will refer to the projective case while polynomials denoted in lower-case, $f(x)$, will refer to the affine case. The source continues to be the thorough [6].

Let C be a curve with equation $F(x, y, z) = 0$ of degree n and let L be a line with the parametric equation

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_1 s + b_1 t \\ a_2 s + b_2 t \\ a_3 s + b_3 t \end{pmatrix},$$

where $\mathbf{a}$ and $\mathbf{b}$ are points on L and s and t varies. To see where L intersects C one can insert the equation of the line into the equation of the curve. This will yield an expression in s and t, $F(\mathbf{a}s + \mathbf{b}t) = 0$. We will have two cases, either $F(\mathbf{a}s + \mathbf{b}t)$ is identically zero and the line L is a component of the curve C, or $F(\mathbf{a}s + \mathbf{b}t)$ is a homogeneous polynomial, also of degree n, with n solutions, not necessarily distinct. The solutions will be ratios $s : t$ where distinct ratios determine distinct points common to C and L. Some solutions may appear more than once, these will be points of multiplicity ≥ 2 and they will be explained further ahead.

Theorem 2.6.1 (Bezout's theorem)**.** *If the equation of a curve C is of degree n, a line either is a component of C or has precisely n points (properly counted) in common with C.*

Theorem 2.6.2. *If C is a curve of order n with no multiple components, then through any point P not on C there pass lines which intersect C in n distinct points.*

A curve of order 1 is a line, a curve of order 2 is called a *conic*, a curve of order 3 a *cubic*, a curve of order 4 a *quartic* etc. and a curve of order n is called a *n-ic*.

Now, let us take a closer look at the points of C and the concept of multiplicity of points. We will proceed in the affine space. Be mindful that some intersections of a curve can occur at points at infinity. Let C have the affine equation $f(x, y) = 0$, $P = (a, b)$ be a point on C and L be a line with parametric equation $x = a + \lambda t$, $y = b + \mu t$ that intersects C in P. The ratio $\lambda : \mu$ will determine the particular line through $P = (a, b)$ on C. As before, we can determine the intersections of C and L by inserting the parametric equation of L into C, $f(a + \lambda t, b + \mu t) = 0$, and examine the roots. As seen before, simply solving $f(a+\lambda t, b+\mu t) = 0$ will yield all intersections between the curve C and the line L. But if we are only interested in how many times L intersect C in the specific point $P = (a, b)$, i.e. P:s multiplicity, then it is sufficient to study the partial derivatives of C. The Taylor expansion of C in $P = (a, b)$ is

$$\begin{aligned} f(a + \lambda t, b + \mu t) =& f(a, b) + \left(f'_x(a, b)\lambda + f'_y(a, b)\mu\right) t + \\ &+ \frac{1}{2!}\left(f''_{xx}(a, b)\lambda^2 + 2f''_{xy}(a, b)\lambda\mu + f''_{yy}(a, b)\mu^2\right) t^2 + \dots \end{aligned}$$

Since the initial condition is that P is a point on the curve C, $f(a, b)$ will always be zero. Suppose that not both $f'_x(a, b)$ and $f'_y(a, b)$ are zero. Then the tangent exists and every line through $P = (a, b)$ has a single intersection with C at P except for the tangent itself where it has a double intersection and $f'_x(a, b)\lambda + f'_y(a, b)\mu = 0$.

This reasoning can be extended to a more general case. Suppose that all partial derivatives up to and including the $r-1$-th in the point $P = (a, b)$ vanish, but not all of the r-th derivatives vanish. Then every line through $P = (a, b)$ will intersect C in $P = (a, b)$ r times except for the tangents. The tangents, of which there will be r, will have more than r intersections and correspond to the roots of

$$\frac{\partial^r f}{\partial x^r}\lambda^r + \binom{r}{1}\frac{\partial^r f}{\partial x^{r-1}\partial y}\lambda^{r-1}\mu + \dots + \binom{r}{r}\frac{\partial^r f}{\partial y^r}\mu^r = 0.$$

P is called an *r-fold point*.

Theorem 2.6.3. *P is an r-fold point of $F(x) = 0$ if and only if all the $(r-1)$-th derivatives of F, but not all r-th derivatives vanish at P.*

A point of multiplicity one is said to be a *simple point*, a point of multiplicity two a *double point*, a point of multiplicity three a *triple point* etc. And, as stated earlier, a point of multiplicity r is called a *r-fold point*. A point of multiplicity two or greater is said to be *singular*. If all tangents at a point are distinct the

point is said to be *ordinary*. A curve with no singular points is called a *non-singular curve* and similarly a curve with one or more singular points is called a *singular curve*. A point not on a given curve is sometimes referred to as a point of multiplicity zero. It is sufficient to establish that all first order derivatives in a given point are zero to determine that the point is singular.

Theorem 2.6.4. *If C is a curve of order n with no multiple components, then through any point P of C of multiplicity r there pass lines which intersect C in $n-r$ distinct points other than P.*

Theorem 2.6.5. *If two curves, of order m and n, have more than mn common points, they have a common component.*

Theorem 2.6.6. *If two curves, of order m and n, have no common components and have multiplicity r_i and s_i at their common points P_i, $i = 1, 2, ...$, then $\sum r_i s_i \leq mn$.*

Theorem 2.6.7. *If two curves of order n intersect in n^2 points, and if exactly mn of these lie on an irreducible curve of order m, then the remaining $n(n-m)$ lie on a curve of order $n-m$.*

Definition 2.6.8. A non-singular point on a curve F is called a *flex* or an *inflection point* if its tangent line has three or more intersections with F.

If a curve has a line component, i.e. a component of degree one, all points on that line component will be flex points. This is called the *degenerate case*. Going forward we will only consider curves without line components.

Definition 2.6.9. Let $f(x_1, \ldots, x_n)$. The *Hessian matrix* is a matrix of the second order partial derivatives of f,

$$H(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Theorem 2.6.10. *The flexes of F are its non-singular points where the determinant of the Hessian matrix*

$$\det(H(x)) = 0.$$

Example 2.6.11. Let $F(x,y,z) = x^3 + y^2z + yz^2 = 0$. Then the Hessian determinant for F is

$$\begin{aligned}\det(H(x,y,z)) &= \begin{vmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} & \frac{\partial^2 F}{\partial x \partial z} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} & \frac{\partial^2 F}{\partial y \partial z} \\ \frac{\partial^2 F}{\partial z \partial x} & \frac{\partial^2 F}{\partial z \partial y} & \frac{\partial^2 F}{\partial z^2} \end{vmatrix} = \\ &= \begin{vmatrix} 6x & 0 & 0 \\ 0 & 2z & 2y+2z \\ 0 & 2y+2z & 2y \end{vmatrix} = \\ &= -24x(y^2+yz+z^2) = 0\end{aligned}$$

The flexes, the points where both F and $\det(H)$ are zero, are $P_1 = (0:1:0)$, $P_2 = (0:0:1)$, $P_3 = (0:1:-1)$, $P_4 = (2:2:-1+i\sqrt{3})$, $P_5 = (2:2:-1-i\sqrt{3})$, $P_6 = (2:-1+i\sqrt{3}:2)$, $P_7 = (2:-1-i\sqrt{3}:2)$, $P_8 = (2:-1+i\sqrt{3}:-1-i\sqrt{3})$ and $P_9 = (2:-1-i\sqrt{3}:-1+i\sqrt{3})$.

Theorem 2.6.12. *Every non-singular curve of order ≥ 3 has at least one flex.*

Theorem 2.6.13. *A non-singular cubic has nine flexes, which have the property that every line joining two of them contains a third.*

Example 2.6.14. As in the previous example, let $F(x,y,z) = x^3+y^2z+yz^2 = 0$. The flexes are the points $P_1, \ldots, P_9$. The twelve distinct lines that the flexes lie on are listed in the table below.

Flexes	**Line**
P_1, P_2, P_3	$L_1 : x = 0$
P_1, P_4, P_9	$L_2 : (-1 + i\sqrt{3})x - 2z = 0$
P_1, P_5, P_8	$L_3 : (-1 - i\sqrt{3})x - 2z = 0$
P_1, P_6, P_7	$L_4 : 2x - 2z = 0$
P_2, P_4, P_5	$L_5 : -2x + 2y = 0$
P_2, P_6, P_8	$L_6 : (1 - i\sqrt{3})x + 2y = 0$
P_2, P_7, P_9	$L_7 : (1 + i\sqrt{3})x + 2y = 0$
P_3, P_4, P_6	$L_8 : (1 + i\sqrt{3})x - 2y - 2z = 0$
P_3, P_5, P_7	$L_9 : (1 - i\sqrt{3})x - 2y - 2z = 0$
P_3, P_8, P_9	$L_{10} : -2x - 2y - 2z = 0$
P_4, P_7, P_8	$L_{11} : 2i\sqrt{3}y + (-3 + i\sqrt{3})z = 0$
P_5, P_6, P_9	$L_{12} : -2i\sqrt{3}y + (-3 - i\sqrt{3})z = 0$

Chapter 3

Elliptic Curves

In this chapter we will explore a particular cubic, the elliptic curve, and a special kind of addition that is associated with it. Point addition on elliptic curves is a way to "add" two points on an elliptic curve to yield another point on the same curve. This is utilised in certain applications of cryptography, which we will explore in chapter 4. We will also look at different representations of elliptic curves. We will start by introducing elliptic curves on Weierstrass form. Later we will look at the Edwards form which have certain advantages related to its addition formula. The sources used in this chapter are [3], [7], [8], [9] and [10].

3.1 Definitions and Addition Law of Elliptic Curves

In this section we will introduce *elliptic curves on Weierstrass form*. This type of curve is the basis on which elliptic curve cryptography rests upon. It is also where the elliptic curve on Edwards form derives from. The source in this section is [3].

Definition 3.1.1. The equation

$$Y^2 = X^3 + AX + B$$

is called a *Weierstrass equation*.

Definition 3.1.2. An *elliptic curve* E is the set of solutions to a Weierstrass equation

$$E : Y^2 = X^3 + AX + B$$

together with a point at infinity, $\mathcal{O}$, and where $4A^3 + 27B^2 \neq 0$.

The condition $4A^3 + 27B^2 \neq 0$ ensures that the curve is non-singular. It comes from the discriminant $\text{Disc}(X^3 + AX + B)$. Theorem 2.3.8 tells us that there is a repeated non-constant factor if the discriminant is zero, and section 2.6 tells us that repeated, non-constant factors yield a singular curve. The point at infinity, $\mathcal{O}$, will be where E and lines parallel to the Y-axis intersect.

Now, we will introduce a new kind of addition that can be applied to ellipic curves. This addition is referred to as *Elliptic curve addition* or *point addition*.

Let P and Q be two points, not necessarily distinct, on an elliptic curve E. A line through P and Q will intersect E in one additional point R, this we know from Bezout's theorem, 2.6.1. Reflecting R across the X-axis yields yet another point on E, namely R'. We denote this relation between P, Q and R' as $P \oplus Q = R'$. Let us look at a first example to illustrate how this works in practice.

Example 3.1.3. Let $E : Y^2 = X^3 + 2X - 3$ be an elliptic curve and let $P = (2, 3)$ and $Q = (1, 0)$ be points on E. The line through P and Q can be described by an equation of the form $Y = kX + m$. Calculating the slope, k, and the constant, m, yields the line $Y = 3X - 3$. The intersections of E and the line $Y = 3X - 3$ can be calculated by inserting $Y = 3X - 3$ into E.

$$(3X - 3)^2 = X^3 + 2X - 3$$

$$X^3 - 9X^2 + 20X - 12 = 0$$

Normally, a cubic equation can be difficult to solve, but in this case we already have two of the solutions, namely $X = 2$ and $X = 1$ from P and Q respectively. So that

$$X^3 - 9X^2 + 20X - 12 = (X - 1)(X - 2)(X - 6).$$

Inserting $X = 6$ into $Y = 3X - 3$ yield $Y = 15$, so that $R = (6, 15)$. All that is left to do is to reflect R across the X-axis. This gives

$$P \oplus Q = R'$$

$$(2, 3) \oplus (1, 0) = (6, -15)$$

The example is illustrated in figure 3.1.

Point addition can also be applied to cases where P and Q are not distinct or one (or both) are the point at infinity. For the next example we shall explore what happens when P and Q are the same point.

Example 3.1.4. Let E and P be as in the previous example. The line that intersects E in P two times is the tangent, see theorem 2.6.3. The gradient of

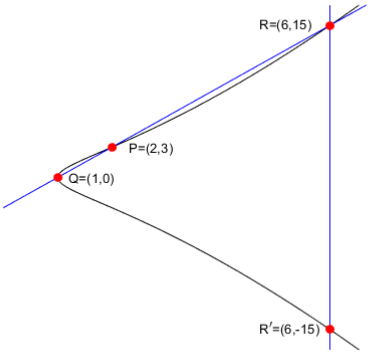

Figure 3.1: $P \oplus Q = R'$ on the curve $E : Y^2 = X^3 + 2X - 3$

E is $\nabla E = (3X^2 + 2, -2Y)$, which is perpendicular to all tangent lines of E. Remember that two perpendicular lines have the relation $k_1 k_2 = -1$, where k_i are their respective slopes. In particular, $\nabla E(P) = \nabla E(2,3) = (14, -6)$ (which yields a slope $k_1 = -3/7$) so that the slope of the tangent at P is $k_2 = 7/3$. The equation for the tangent at P is $Y = 7X/3 - 5/3$. The intersection between E and the tangent at P is calculated in the same way as in the previous example.

$$\left(\frac{7}{3}X - \frac{5}{3}\right)^2 = X^3 + 2X - 3$$

$$9X^3 - 49X^2 + 88X - 52 = (X-2)^2(9X - 13) = 0$$

Inserting $X = 13/9$ into the equation for the tangent yields $Y = 46/27$ so that the third intersection of E and the tangent is at $R = \left(\frac{13}{9}, \frac{46}{27}\right)$. Finally, reflecting R across the X-axis yields

$$R' = P \oplus P = \left(\frac{13}{9}, -\frac{46}{27}\right).$$

The example is illustrated in figure 3.2.

There are two more cases to consider. The first is when P and Q are each others reflection points, that is when $Q = P'$.

Example 3.1.5. Let $E : Y^2 = X^3 + 2X - 3$, $P = (2,3)$ and let $P' = (2,-3)$ be P:s reflection point. First, observe that P' does lie on E (and so does all

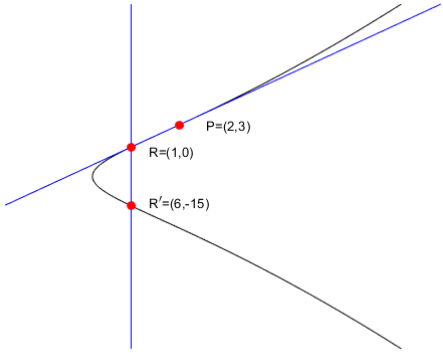

Figure 3.2: $P \oplus P = R'$ on the curve $E : Y^2 = X^3 + 2X - 3$

the reflection points to any point on the curve). Since P and P' share the same X-coordinate they lie on the line $X = 2$. With what we know of intersections of curves from section 2.6, and Bezout's theorem 2.6.1 in particular, we can conclude that the third point of intersection of E and $X = 2$ is at the point at infinity. Reflecting the point at infinity across the X-axis yields, again, the point at infinity. We conclude that $P \oplus P' = \mathcal{O}$. See figure 3.3.

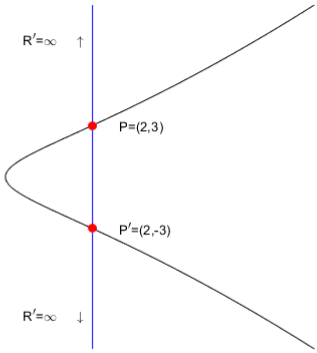

Figure 3.3: $P \oplus P' = R'$ on the curve $E : Y^2 = X^3 + 2X - 3$

The final case is when one of the points is the point at infinity.

Example 3.1.6. Let $E : Y^2 = X^3+2X-3$ and $P = (2,3)$. The line connecting P and the point at infinity is $X = 2$. As in the previous example, $X = 2$ intersects E in $P' = (2,-3)$. Reflecting P' across the X-axis yields P. So $P \oplus \mathcal{O} = P$.

To summarise these examples we have the following two theorems.

Theorem 3.1.7. *Let E be an elliptic curve. Then the addition law on E has the following properties:*

1.	$(P+Q)+R = P+(Q+R)$	*for all* $P,Q,R \in E$.	*[Associative]*
2.	$P+\mathcal{O} = \mathcal{O}+P = P$	*for all* $P \in E$.	*[Identity]*
3.	$P+(-P) = \mathcal{O}$	*for all* $P \in E$.	*[Inverse]*
4.	$P+Q = Q+P$	*for all* $P,Q \in E$.	*[Commutative]*

That is, E and the addition law forms an abelian group.

The second theorem summarises the computations.

Theorem 3.1.8 (Elliptic Curve Addition Algorithm)**.** *Let*

$$E : Y^2 = X^3 + AX + B$$

be an elliptic curve and let P_1 and P_2 be points on E.

1. *If $P_1 = \mathcal{O}$, then $P_1 + P_2 = P_2$.*
2. *Otherwise, if $P_2 = \mathcal{O}$, then $P_1 + P_2 = P_1$.*
3. *Otherwise, write $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$.*
4. *If $x_1 = x_2$ and $y_1 = -y_2$, then $P_1 + P_2 = \mathcal{O}$.*
5. *Otherwise, define λ by*

$$\lambda = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1} & \text{if } P_1 \neq P_2, \\[2ex] \dfrac{3x_1^2 + A}{2y_1} & \text{if } P_1 = P_2, \end{cases}$$

and let

$$x_3 = \lambda^2 - x_1 - x_2 \quad \text{and} \quad y_3 = \lambda(x_1 - x_3) - y_1.$$

Then $P_1 + P_2 = (x_3, y_3)$.

Example 3.1.9 (Deduction of the Elliptic Curve Algorithm)**.** Let $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ be two points, not necessarily distinct, on an elliptic curve $E : Y^2 = X^3 + AX + B$, and let $L = PQ$ be the line through P and Q. We wish to find the coordinates of the point $R' = P \oplus Q$. In order to do so, we must first find the coordinates of the point R. We know from Bezout's theorem, 2.6.1, that L will intersect E in the additional point $R = (x_3, y_3)$. In order to determine R we use the line equation $Y = \lambda X + m$. As the line equation and calculating intersections between curves should be familiar, we will keep the explanations brief. First of all, there is a small difference in how L is calculated depending on if $P \neq Q$ or if $P = Q$. If P and Q are distinct then the slope, λ, is determined by

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1}.$$

If $P = Q$, then L is the tangent to E on the point P and λ is determined using the gradient, $\nabla E(P) = \nabla E(x_1, y_1) = (3x_1^2 + A, -2y_1)$. Then

$$\lambda = -\frac{1}{\frac{-2y_1}{3x_1^2 + A}} = \frac{3x_1^2 + A}{2y_1},$$

since the gradient yields the slope perpendicular to the tangent. So,

$$L : Y = \lambda X + m \quad \text{where} \quad \lambda = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1} & \text{if } P \neq Q, \\ \dfrac{3x_1^2 + A}{2y_1} & \text{if } P = Q. \end{cases}$$

Inserting the point P into the equation yields the value of m.

$$y_1 = \lambda x_1 + m \iff m = y_1 - \lambda x_1$$

With these values for λ and m, $L = PQ$ has the equation

$$L : Y = \lambda X + y_1 - \lambda x_1 = \lambda(X - x_1) + y_1.$$

Now, the intersection between L and E is calculated as follows:

$$(\lambda(X - x_1) + y_1)^2 = X^3 + AX + B \iff$$

$$X^3 - \lambda^2 X^2 + (A + 2\lambda^2 x_1 - 2\lambda y_1)X + B - \lambda^2 x_1^2 + 2\lambda x_1 y_1 - y_1^2 = 0$$

Since we already know that P and Q are points on E, finding the third point, R, is easy. By dividing the equation stated above by $X - x_1$ and by $X - x_2$ or by $(X - x_1)^2$ respectively, we see that

$$(X - x_1)(X - x_2)(X + x_1 + x_2 - \lambda^2) = 0 \quad \text{resp.}$$

$$(X - x_1)^2(X + 2x_1 - \lambda^2) = 0$$

Note that these equations are equal if $x_1 = x_2$.

Remark. If the reader wishes to do these calculations themselves, they will notice that there will be some remainders when dividing the equation by $X - x_1$ and $X - x_2$ or $(X - x_1)^2$ respectively. These remainders are equal to zero since we know that P and Q are points on E. But if the reader wishes, they can also expand the remainders and use the appropriate value of λ and the fact that $x_1^3 + Ax_1 + B - y_1^2 = 0$ and $x_2^3 + Ax_2 + B - y_2^2 = 0$ to explicitly show that these remainders are equal to zero.

So the coordinates for the point R are

$$(x_3, y_3) = (\lambda^2 - x_1 - x_2, \lambda(x_3 - x_1) + y_1) \quad \text{resp.}$$

$$(x_3, y_3) = (\lambda^2 - 2x_1, \lambda(x_3 - x_1) + y_1).$$

Again, the coordinates above are equal if $x_1 = x_2$.

Now all that remains is to reflect the Y-coordinate across the X-axis. So the coordinates for $R' = P \oplus Q$ are

$$R' = (x', y') = (\lambda^2 - x_1 - x_2, \lambda(x_1 - x_3) - y_1)$$

Let us return to our previous examples, this time using the algorithm.

Example 3.1.10. Let $E : Y^2 = X^3 + 2X - 3$, $P = (2, 3)$ and $Q = (1, 0)$. First, let us look at $P \oplus Q$. P and Q are distinct and neither of them is a point at infinity so let us follow the addition algorithm from step 3. $P_1 = P$ and $P_2 = Q$ so that $x_1 = 2$, $y_1 = 3$, $x_2 = 1$ and $y_2 = 0$ which gives

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1} = \frac{0 - 3}{1 - 2} = 3.$$

And the values for x_3 and y_3 are

$$x_3 = \lambda^2 - x_1 - x_2 = 3^2 - 1 - 2 = 6$$

$$y_3 = \lambda(x_1 - x_3) - y_1 = 3(1 - 6) - 0 = -15$$

so that

$$R' = P \oplus Q = (6, -15).$$

Now, lets look at $P \oplus P$ and follow the algorithm from step 3 for $P_1 = P_2$. The

only significant difference in the calculations compared to $P_1 \neq P_2$ is that of the value for λ.

$$\lambda = \frac{3x_1^2 + A}{2y_1} = \frac{3 \cdot 2^2 + 2}{2 \cdot 3} = \frac{7}{3}$$

As for $P_1 \neq P_2$, the values of x_3 and y_3 are given by

$$x_3 = \lambda^2 - x_1 - x_2 = \left(\frac{7}{3}\right)^2 - 2 \cdot 2 = \frac{13}{9}$$

$$y_3 = \lambda(x_1 - x_3) - y_1 = \frac{7}{3}\left(2 - \frac{13}{9}\right) - 3 = -\frac{46}{27}$$

so that

$$R' = P \oplus P = \left(\frac{13}{9}, -\frac{46}{27}\right).$$

3.2 Elliptic Curves over Finite Fields

In the previous section we looked at the Weierstrass curve in a general algebraically closed field. We will now move on to the finite field. It is curves on finite fields that are utilised when working with the cryptographic applications, as we will see in chapter 4. The source used in this section is [3].

Definition 3.2.1. Let $p \geq 3$ be a prime. An *elliptic curve over* $\mathbb{F}_p$ is an equation of the form

$$E : Y^2 = X^3 + AX + B \qquad \text{with } A, B \in \mathbb{F}_p \text{ satisfying } 4A^3 + 27B^2 \neq 0.$$

The *set of points on E with coordinates in* $\mathbb{F}_p$ is the set

$$E(\mathbb{F}_p) = \{(x, y) : x, y \in \mathbb{F}_p \text{ satisfy } y^2 = x^3 + Ax + B\} \cup \{\mathcal{O}\}.$$

It is possible to use point addition on curves over $\mathbb{F}_{2^k}$, $k \in \mathbb{Z}_+$ but it requires some additional conditions and will not be covered in this essay.

Theorem 3.2.2. *Let E be an elliptic curve over $\mathbb{F}_p$ and let P and Q be points in $E(\mathbb{F}_p)$.*

1. *The elliptic curve addition algorithm 3.1.8 applied to P and Q yields a point in $E(\mathbb{F}_p)$. We denote this point by $P + Q$.*
2. *The addition law on $E(\mathbb{F}_p)$ satisfies all of the properties in 3.1.7. In other words, this addition law makes $E(\mathbb{F}_p)$ into a finite abelian group.*

To understand the following examples it can be beneficial to refresh some knowledge about number theory, particularly modulus calculation and its inverses. In these examples we will again look at the curve $E : Y^2 = X^3 + 2X - 3$ but for a specific finite field. Notice that the quantity $4A^3 + 27B^2 = 275 = 5^2 11$ for this curve. This means that the quantity will be equal to 0 in $\mathbb{F}_5$ and $\mathbb{F}_{11}$ and thus does not meet the criteria of definition 3.2.1.

Example 3.2.3. Let us look at $E : Y^2 = X^3 + 2X - 3$ over the finite field $\mathbb{F}_7$. To find the points on E, we insert each value $X \pmod 7$ into E to see whether or not the result is a square modulo 7.

$X = a$	$X^3 + 2X - 3 = b$	$Y^2 \equiv b \pmod 7$?	Solutions?
$X = 0$	$b = -3$	$Y^2 \equiv -3 \equiv 4 \pmod 7$	$Y \equiv 2 \pmod 7$ and $Y \equiv 5 \pmod 7$
$X = 1$	$b = 0$	$Y^2 \equiv 0 \pmod 7$	$Y = 7n$, $n \in \mathbb{Z}$
$X = 2$	$b = 9$	$Y^2 \equiv 9 \equiv 2 \pmod 7$	$Y \equiv 3 \pmod 7$ and $Y \equiv 4 \pmod 7$
$X = 3$	$b = 30$	$Y^2 \equiv 30 \equiv 2 \pmod 7$	$Y \equiv 3 \pmod 7$ and $Y \equiv 4 \pmod 7$
$X = 4$	$b = 69$	$Y^2 \not\equiv 69 \equiv 6 \pmod 7$	No solutions
$X = 5$	$b = 132$	$Y^2 \not\equiv 132 \equiv 6 \pmod 7$	No solutions
$X = 6$	$b = 225$	$Y^2 \equiv 225 \equiv 1 \pmod 7$	$Y \equiv 1 \pmod 7$ and $Y \equiv 6 \pmod 7$

Table 3.1: Quadratic residues and non-residues modulo 7.

The table 3.1 indicates that E has 10 points in $\mathbb{F}_7$, namely

$$E(\mathbb{F}_7) = \{\mathcal{O}, (0,2), (0,5), (1,0), (2,3), (2,4), (3,3), (3,4), (6,1), (6,6)\}.$$

Example 3.2.4. Let us use the addition algorithm, 3.1.8, on $E : Y^2 = X^3 + 2X - 3$ in F_7. Let $P = (2,3)$ and let $Q = (6,6)$. Neither P nor Q is the point at infinity, $\mathcal{O}$, and the points have distinct X- and Y-coordinates. Therefore, we calculate $P + Q = (2,3) + (6,6)$ by first finding the value for λ, where

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1} = \frac{6-3}{6-2} = \frac{3}{4} = 3 \cdot \frac{1}{4} = 3 \cdot 2 = 6.$$

Remark. Fractions do not function the same way they do in calculus. $\frac{1}{4}$ is to be understood as the inverse of 4 in the current finite field, which, in this case, can be found by solving $4a \equiv 1 \pmod 7$, which yield $a \equiv 2 \pmod 7$.

Continuing,

$$\begin{aligned} x_3 =& \lambda^2 - x_1 - x_2 = 6^2 - 2 - 6 = 28 = 0 \\ y_3 =& \lambda(x_1 - x_3) - y_1 = 6(2 - 0) - 3 = 9 = 2. \end{aligned}$$

Notice that the last equalities for both x_3 and y_3, $28 = 0$ resp. $9 = 2$, are true because we are in the field $\mathbb{F}_7$ and $28 \equiv 0 \pmod 7$ and $9 \equiv 2 \pmod 7$. We conclude that

$$(2, 3) + (6, 6) = (0, 2)$$

in $\mathbb{F}_7$.

Example 3.2.5. Let $E : Y^2 = X^3 + 2X - 3$ in $\mathbb{F}_7$ and let $P = (2, 3)$. What is $P + P$? Recall,

$$\lambda = \frac{3x_1^2 + A}{2y_1} = \frac{3 \cdot 2^2 + 2}{2 \cdot 3} = \frac{14}{6} = \frac{7}{3} = 0$$

As in the previous example, the equality holds because we are in $\mathbb{F}_7$ and $7 = 0$. Now,

$$\begin{aligned} x_3 =& \lambda^2 - 2x_1 = 0^2 - 2 \cdot 2 = -4 = 3 \\ y_3 =& \lambda(x_1 - x_3) - y_1 = 0(2 - 3) - 3 = -3 = 4 \end{aligned}$$

so that

$$(2, 3) + (2, 3) = (3, 4).$$

The complete addition table for $E(\mathbb{F}_7) : Y^2 = X^3 + 2X - 3$ can be found in table 3.2.

For comparison, the addition table for $E : Y^2 = X^3 + 2X - 3$ in $\mathbb{F}_3$ can be found in table 3.3.

3.3 Edwards Curve

An issue when utilising point addition on an elliptic curve on Weierstrass form is that all curves contain a point at infinity. It is usually not a problem as long as the computations are done by hand but transferring the addition algorithm to computers can cause problems. This is because "infinity" on computers

+	$\mathcal{O}$	$(0,2)$	$(0,5)$	$(1,0)$	$(2,3)$	$(2,4)$	$(3,3)$	$(3,4)$	$(6,1)$	$(6,6)$
$\mathcal{O}$	$\mathcal{O}$	$(0,2)$	$(0,5)$	$(1,0)$	$(2,3)$	$(2,4)$	$(3,3)$	$(3,4)$	$(6,1)$	$(6,6)$
$(0,2)$	$(0,2)$	$(2,4)$	$\mathcal{O}$	$(3,4)$	$(0,5)$	$(6,6)$	$(1,0)$	$(6,1)$	$(2,3)$	$(3,3)$
$(0,5)$	$(0,5)$	$\mathcal{O}$	$(2,3)$	$(3,3)$	$(6,1)$	$(0,2)$	$(6,6)$	$(1,0)$	$(3,4)$	$(2,4)$
$(1,0)$	$(1,0)$	$(3,4)$	$(3,3)$	$\mathcal{O}$	$(6,6)$	$(6,1)$	$(0,5)$	$(0,2)$	$(2,4)$	$(2,3)$
$(2,3)$	$(2,3)$	$(0,5)$	$(6,1)$	$(6,6)$	$(3,4)$	$\mathcal{O}$	$(2,4)$	$(3,3)$	$(1,0)$	$(0,2)$
$(2,4)$	$(2,4)$	$(6,6)$	$(0,2)$	$(6,1)$	$\mathcal{O}$	$(3,3)$	$(3,4)$	$(2,3)$	$(0,5)$	$(1,0)$
$(3,3)$	$(3,3)$	$(1,0)$	$(6,6)$	$(0,5)$	$(2,4)$	$(3,4)$	$(2,3)$	$\mathcal{O}$	$(0,2)$	$(6,1)$
$(3,4)$	$(3,4)$	$(6,1)$	$(1,0)$	$(0,2)$	$(3,3)$	$(2,3)$	$\mathcal{O}$	$(2,4)$	$(6,6)$	$(0,5)$
$(6,1)$	$(6,1)$	$(2,3)$	$(3,4)$	$(2,4)$	$(1,0)$	$(0,5)$	$(0,2)$	$(6,6)$	$(3,3)$	$\mathcal{O}$
$(6,6)$	$(6,6)$	$(3,3)$	$(2,4)$	$(2,3)$	$(0,2)$	$(1,0)$	$(6,1)$	$(0,5)$	$\mathcal{O}$	$(3,4)$

Table 3.2: Addition table for $E : Y^2 = X^3 + 2X - 3$ in $\mathbb{F}_7$.

+	$\mathcal{O}$	$(0,0)$	$(1,0)$	$(2,0)$
$\mathcal{O}$	$\mathcal{O}$	$(0,0)$	$(1,0)$	$(2,0)$
$(0,0)$	$(0,0)$	$\mathcal{O}$	$(2,0)$	$(1,0)$
$(1,0)$	$(1,0)$	$(2,0)$	$\mathcal{O}$	$(0,0)$
$(2,0)$	$(2,0)$	$(1,0)$	$(0,0)$	$\mathcal{O}$

Table 3.3: Addition table for $E : Y^2 = X^3 + 2X - 3$ in $\mathbb{F}_3$.

is not actually infinity, rather a very large fixed number, and computations are done numerically. Another aspect is that point addition on a curve on Weierstrass form utilises different cases depending on the points being added. By transforming elliptic curves on Weierstrass form into another form both these issues can largely be avoided. We will first define this other form, the Edwards curve, of elliptic curves and then prove that it is another representation of elliptic curves on Weierstrass form. The sources in this section are the original paper on Edwards curve, [7], and papers building on its findings, [8] [9], and [10].

Definition 3.3.1 (Twisted Edwards curve and Edwards curve)**.** Fix a field K with $\mathrm{char}(K) \neq 2$. Fix distinct non-zero elements $a, d \in K$. The *Twisted Edwards curve* with coefficients a and d is the curve

$$E_{E,a,d} : ax^2 + y^2 = 1 + dx^2y^2.$$

An *Edwards curve* is a twisted Edwards curve with $a = 1$, denoted

$$E_{E,d} : x^2 + y^2 = 1 + dx^2y^2.$$

Remark. Note that although we will use the Twisted Edwards curve in order to prove the transformations between different representations of elliptic curves, we will mainly utilise the non-twisted version for our application later on.

To see that an elliptic curve on Edwards form is equivalent to its Weierstrass counterpart we shall look at the algebraic transformation between them. To do so we will use a another form, called a Montgomery curve, as stepping stone between Edwards and Weierstrass. This is solely to make the computations easier and the Montgomery curve will not be explored further in this essay. Though, it is worth knowing that Montgomery curves have their own place in cryptography. [10] presents some of its benefits.

Definition 3.3.2 (Montgomery curve)**.** Fix a field K with $\text{char}(K) \neq 2$. Fix $\alpha \in K \setminus \{2, -2\}$ and $\beta \in K \setminus \{0\}$. The *Montgomery curve* with coefficients α and β is the curve

$$E_{M,\alpha,\beta} : \beta v^2 = u^3 + \alpha u^2 + u.$$

Remark. In order to easier distinguish between the different forms used in this essay, the Weierstrass form will always use upper-case letters X, Y, A, B, the Edwards form lower-case letters x, y, a, d and the Montgomery form will use u, v, α, β.

With the different forms of elliptic curves defined we shall start looking at the algebraic transformations between them. First, a short explanation on how all the tables displaying the transformations are to be understood. The tables display only one direction of the transformation, therefore, each pair of forms have two tables associated with them. This is because it is not simple to discern the inverse transformation from its original transformation.

The contents of the tables are to be understood as follows: The first two rows display the equations of the different elliptic curve forms and their respective conditions for existence. The third row displays the birational map from a point of the first form to a point of the second form. The final row displays the relations of the constants of the two forms.

In practice, this means that we use the map of the third row to transform a particular point, P_1, on curve E_1 of the first form into its counterpart, P_2, on curve E_2 of the second form. Similarly, we use the fourth row in order to transform the curve E_1 of the first form into the curve E_2 of the second form.

However, in order to prove that these transformations are sound one inserts the mapping of "E_1 to E_2" into E_2 in order to yield E_1. This might seem confusing but the deductions will hopefully make it clear. Henceforth, we will use the notation $E_{W,A,B}$ for a curve on Weierstrass form with constants A and B.

Example 3.3.3 (Transformation of Montgomery into Weierstrass)**.** To verify the transformation of a curve of Montgomery form into a curve of Weierstrass

Montgomery	**Weierstrass**
$E_{M,\alpha,\beta} : \beta v^2 = u^3 + \alpha u^2 + u$	$E_{W,A,B} : Y^2 = X^3 + AX + B$
$\beta(\alpha^2 - 4) \neq 0$	$4A^3 + 27B^2 \neq 0$
$(u, v) \longmapsto (X, Y) = \left(\frac{u + \frac{\alpha}{3}}{\beta}, \frac{v}{\beta}\right)$	
$A = \frac{1}{\beta^2} - \frac{\alpha^2}{3\beta^2}$,	$B = \frac{\alpha(2\alpha^2 - 9)}{27\beta^3}$

Table 3.4: Transformation from Montgomery into Weierstrass.

form, insert

$$(X, Y) = \left(\frac{u + \frac{\alpha}{3}}{\beta}, \frac{v}{\beta}\right), A = \frac{1}{\beta^2} - \frac{\alpha^2}{3\beta^2} \text{ and } B = \frac{\alpha(2\alpha^2 - 9)}{27\beta^3}$$

into $Y^2 = X^3 + AX + B$ in order to yield $\beta v^2 = u^3 + \alpha u^2 + u$, see table 3.4. The calculations are not difficult and the explanations will be kept brief.

$$\left(\frac{v}{\beta}\right)^2 = \left(\frac{u + \frac{\alpha}{3}}{\beta}\right)^3 + \left(\frac{1}{\beta^2} - \frac{\alpha^2}{3\beta^2}\right)\left(\frac{u + \frac{\alpha}{3}}{\beta}\right) + \frac{\alpha(2\alpha^2 - 9)}{27\beta^3}$$

Multiply both sides by β^3.

$$\begin{aligned}
\beta v^2 &= \left(u + \frac{\alpha}{3}\right)^3 + \left(1 - \frac{\alpha^2}{3}\right)\left(u + \frac{\alpha}{3}\right) + \frac{\alpha(2\alpha^2 - 3^2)}{3^3} = \\
&= u^3 + \alpha u^2 + \frac{\alpha^2 u}{3} + \frac{\alpha^3}{3^3} + u + \frac{\alpha}{3} - \frac{\alpha^2 u}{3} - \frac{\alpha^3}{3^2} + \frac{2\alpha^3}{3^3} - \frac{\alpha}{3} = \\
&= u^3 + \alpha u^2 + u + \frac{3\alpha^3}{3^3} - \frac{\alpha^3}{3^2}
\end{aligned}$$

$$\beta v^2 = u^3 + \alpha u^2 + u$$

Thus we have shown that the transformation between Montgomery and Weierstrass is sound.

Twisted Edwards	**Montgomery**
$E_{E,a,d} : ax^2 + y^2 = 1 + dx^2y^2$	$E_{M,\alpha,\beta} : \beta v^2 = u^3 + \alpha u^2 + u$
$a \neq 0,\ d \neq 0,\ a \neq d$	$\beta(\alpha^2 - 4) \neq 0$
$(x, y) \longmapsto (u, v) = \left(\frac{1+y}{1-y}, \frac{1+y}{x(1-y)}\right)$	
$\alpha = \frac{2(a+d)}{a-d}, \qquad \beta = \frac{4}{a-d}$	

Table 3.5: Transformation from Twisted Edwards into Montgomery.

Next, we transform a curve of Twisted Edwards form into a curve of Montgomery form.

Example 3.3.4 (Transformation of Twisted Edwards into Montgomery)**.** Similar to the previous example the idea is to insert

$$(u, v) = \left(\frac{1+y}{1-y}, \frac{1+y}{x(1-y)}\right),\ \alpha = \frac{2(a+d)}{a-d} \text{ and } \beta = \frac{4}{a-d}$$

into $\beta v^2 = u^3 + \alpha u^2 + u$ in order to yield $ax^2 + y^2 = 1 + dx^2y^2$. See table 3.5.

$$\frac{4}{a-d}\left(\frac{1+y}{x(1-y)}\right)^2 = \left(\frac{1+y}{1-y}\right)^3 + \frac{2(a+d)}{a-d}\left(\frac{1+y}{1-y}\right)^2 + \frac{1+y}{1-y}$$

Multiply both sides with $\frac{x^2(a-d)(1-y)^3}{1+y}$. This is allowed because $1 + y \neq 0$. If $y = -1$ then $a = d$ in Twisted Edwards form, which is a violation of its

criteria.

$$4(1+y)(1-y) = x^2\left((a-d)(1+y)^2 + 2(a+d)(1+y)(1-y) + \right.$$
$$\left. +(a-d)(1-y)^2\right)$$

$$4(1-y^2) = x^2\left((a-d)(1+2y+y^2) + 2(a+d)(1-y^2) + \right.$$
$$\left. +(a-d)(1-2y+y^2)\right) =$$
$$= x^2\left((a-d-2a-2d+a-d)y^2 + (2a-2d-2a+2d)y^1 + \right.$$
$$\left. (a-d+2a+2d+a-d)y^0\right)$$

$$4(1-y^2) = x^2(-4dy^2 4a)$$

$$ax^2 + y^2 = 1 + dx^2y^2$$

And so, the transformation between Twisted Edwards and Montgomery is also proven sound.

It should now be obvious that we can construct a transformation between Twisted Edwards and Weierstrass since we can construct the transformations from Montgomery to Weierstrass and from Twisted Edwards to Montgomery. The calculations for transforming Twisted Edwards into Weierstrass are no more difficult than the previous two transformations, though the expressions will be more voluminous.

Remark. As the reader may notice, the expressions in table 3.6 could be simplified into much shorter expressions. It is a deliberate choice not to do so in order to highlight how the transformations between Montgomery and Weierstrass respectively Twisted Edwards and Montgomery have been merged.

Example 3.3.5 (Transformation of Twisted Edwards into Weierstrass)**.** As in the previous examples, we insert the new variables and constants into the

Twisted Edwards	**Weierstrass**
$E_{E,a,d} : ax^2 + y^2 = 1 + dx^2y^2$	$E_{W,A,B} : Y^2 = X^3 + AX + B$
$a \neq 0,\ d \neq 0,\ a \neq d$	$4\alpha^3 + 27\beta^2 \neq 0$

$$(x, y) \longmapsto (X, Y) = \left(\frac{\dfrac{1+y}{1-y} + \dfrac{\dfrac{2(a+d)}{(a-d)}}{3}}{\dfrac{4}{a-d}}, \dfrac{\dfrac{1+y}{x(1-y)}}{\dfrac{4}{a-d}} \right)$$

$$A = \frac{1}{\left(\dfrac{4}{a-d}\right)^2} - \frac{\left(\dfrac{2(a+d)}{a-d}\right)^2}{3\left(\dfrac{4}{a-d}\right)^2}, \qquad B = \frac{\dfrac{2(a+d)}{a-d}\left(2\left(\dfrac{2(a+d)}{a-d}\right)^2 - 9\right)}{27\left(\dfrac{4}{a-d}\right)^3}$$

Table 3.6: Transformation from Twisted Edwards into Weierstrass.

Weierstrass form in order to yield the Twisted Edwards form, see table 3.6.

$$\left(\frac{\dfrac{1+y}{x(1-y)}}{\dfrac{4}{a-d}} \right)^2 = \left(\frac{\dfrac{1+y}{1-y} + \dfrac{\dfrac{2(a+d)}{(a-d)}}{3}}{\dfrac{4}{a-d}} \right)^3 +$$

$$+ \left(\frac{1}{\left(\dfrac{4}{a-d}\right)^2} - \frac{\left(\dfrac{2(a+d)}{a-d}\right)^2}{3\left(\dfrac{4}{a-d}\right)^2} \right) \left(\frac{\dfrac{1+y}{1-y} + \dfrac{\dfrac{2(a+d)}{(a-d)}}{3}}{\dfrac{4}{a-d}} \right) +$$

$$+ \frac{\dfrac{2(a+d)}{a-d}\left(2\left(\dfrac{2(a+d)}{a-d}\right)^2 - 9\right)}{27\left(\dfrac{4}{a-d}\right)^3}$$

For abbreviation, let $k_1 = a + d$, $k_2 = a - d$, $y_1 = 1 + y$ and $y_2 = 1 - y$.

$$\left(\frac{\frac{y_1}{xy_2}}{\frac{2^2}{k_2}}\right)^2 = \left(\frac{\frac{y_1}{y_2} + \frac{\frac{2k_1}{k_2}}{3}}{\frac{2^2}{k_2}}\right)^3 + \left(\frac{1}{\left(\frac{2^2}{k_2}\right)^2} - \frac{\left(\frac{2k_1}{k_2}\right)^2}{3\left(\frac{2^2}{k_2}\right)^2}\right)\left(\frac{\frac{y_1}{y_2} + \frac{\frac{2k_1}{k_2}}{3}}{\frac{2^2}{k_2}}\right) + \\ + \frac{\frac{2k_1}{k_2}\left(2\left(\frac{2k_1}{k_2}\right)^2 - 3^2\right)}{3^3\left(\frac{2^2}{k_2}\right)^3}$$

Simplify by factoring out denominators.

$$\frac{k_2^2 y_1^2}{2^4 x^2 y_2^2} = \frac{k_2^3}{2^6}\left(\frac{y_1}{y_2} + \frac{2k_1}{3k_2}\right)^3 + \frac{k_2^3}{2^6 3}\left(3 - \frac{2^2 k_1^2}{k_2^2}\right)\left(\frac{y_1}{y_2} + \frac{2k_1}{3k_2}\right) + \\ + \frac{k_2^3}{2^6 3^3}\left(\frac{2^4 k_1^3}{k_2^3} - \frac{2 \cdot 3^2 k_1}{k_2}\right)$$

Multiply both sides with $2^6 3^3 x^2 y_2^3$ in order to eliminate the denominators.

$$2^2 3^3 k_2^2 y_1^2 y_2 = x^2\left(3^3 k_2^3 y_2^3\left(\frac{y_1}{y_2} + \frac{2k_1}{3k_2}\right)^3 + 3^2 k_2^3 y_2^3\left(3 - \frac{2^2 k_1^2}{k_2^2}\right)\left(\frac{y_1}{y_2} + \frac{2k_1}{3k_2}\right) + \right. \\ \left. + k_2^3 y_2^3\left(\frac{2^4 k_1^3}{k_2^3} - \frac{2 \cdot 3^2 k_1}{k_2}\right)\right)$$

$$2^2 3^3 k_2^2 y_1^2 y_2 = x^2\left((3k_2 y_1 + 2k_1 y_2)^3 + (3k_2^2 - 2^2 k_1^2)(3^2 k_2 y_1 y_2^2 + 2 \cdot 3 k_1 y_2^3) + \right. \\ \left. + 2^4 k_1^3 y_2^3 - 2 \cdot 3^2 k_1 k_2^2 y_2^3\right)$$

Expand the parentheses.

$$2^2 3^3 k_2^2 y_1^2 y_2 = x^2\left(3^3 k_2^3 y_1^3 + 2 \cdot 3^3 k_1 k_2^2 y_1^2 y_2 + 2^2 3^2 k_1^2 k_2 y_1 y_2^2 + 2^3 k_1^3 y_2^3 + \right. \\ + 3^3 k_2^3 y_1 y_2^2 + 2 \cdot 3^2 k_1 k_2^2 y_2^3 - 2^2 3^2 k_1^2 k_2 y_1 y_2^2 - \\ \left. - 2^3 3 k_1^3 y_2^3 + 2^4 k_1^3 y_2^3 - 2 \cdot 3^2 k_1 k_2^2 y_2^3\right)$$

Gather all coefficients with their respective variables.

$$2^2 3^3 k_2^2 y_1^2 y_2 = x^2\left(3^3 k_2^3 y_1^3 + (2^3 - 2^3 3 + 2^4) k_1^3 y_2^3 + (2^2 3^2 - 2^2 3^2) k_1^2 k_2 y_1 y_2^2 + \right. \\ + (2 \cdot 3^2 - 2 \cdot 3^2) k_1 k_2^2 y_2^3 + 2 \cdot 3^3 k_1 k_2^2 y_1^2 y_2 + \\ \left. + 3^3 k_2^3 y_1 y_2^2\right)$$

Simplify.

$$2^2 3^3 k_2^2 y_1^2 y_2 = x^2 \left(3^3 k_2^3 y_1^3 + 2 \cdot 3^3 k_1 k_2^2 y_1^2 y_2 + 3^3 k_2^3 y_1 y_2^2\right)$$

Multiply both sides with $\frac{1}{3^3 k_2^2 y_1}$. Similarly to the transformation between the Montgomery form and the Twisted Edwards form, this is in order because $k_2^2 y_1 = (a-d)^2(1+y) \neq 0$. Either factor being equal to 0 would be a violation of the Twisted Edwards form criteria.

$$2^2 y_1 y_2 = x^2(k_2 y_1^2 + 2k_1 y_1 y_2 + k_2 y_2^2)$$

Now it is time to change back into our original variables and constants. Remember, $k_1 = a + d$, $k_2 = a - d$, $y_1 = 1 + y$ and $y_2 = 1 - y$.

$$\begin{aligned} 4(1+y)(1-y) &= x^2 \left((a-d)(1+y)^2 + 2(a+d)(1+y)(1-y) + \right. \\ &\qquad \left. + (a-d)(1-y)^2\right) = \\ &= x^2 \left((2a - 2a - 4d)y^2 + (2a - 2a + 2d - 2d)y^1 + \right. \\ &\qquad \left. + (4a + 2d - 2d)y^0\right) = \\ &= x^2(-4dy^2 + 4a) \end{aligned}$$

Expand both the left and right hand sides and rearrange the equation.

$$1 - y^2 = ax^2 - dx^2 y^2 \quad \Longleftrightarrow \quad ax^2 + y^2 = 1 + dx^2 y^2$$

Thus, the transformation between Twisted Edwards form and Weierstrass form is proven sound.

We have now shown all the transformations in one direction. The inverse transformations are listed in tables 3.7, 3.8, 3.9 and 3.10. The process of proving these transformations are similar to those already shown and will therefore be omitted. However, some of the transformations require a few more conditions and those will be touched upon briefly. The transformations from Weierstrass into Montgomery, table 3.7, Weierstrass into Twisted Edwards, table 3.9, and Weierstrass into Edwards, table 3.10, have two extra conditions. K is a field which can be finite or non-finite. A *quadratic residue* is a concept from number theory. In short, q is a quadratic residue modulo n if there exist an x such that $x^2 \equiv q \pmod{n}$. If the reader wish to learn more, any book on basic number theory will be sufficient.

It is possible to use table 3.6 to transform a non-twisted Edwards curve into a Weierstrass curve simply by setting $a = 1$. However, we can not use this

Weierstrass	**Montgomery**
$E_{W,A,B} : Y^2 = X^3 + AX + B$	$E_{M,\alpha,\beta} : \beta v^2 = u^3 + \alpha u + u$
$4A^3 + 27B^2 \neq 0$	$\beta(\alpha^2 - 4) \neq 0$

1. $z^3 + Az + B = 0$ has one root $\varphi \in K$; and
2. $3\varphi^2 + A$ is a quadratic residue in K.

$$(X, Y) \longmapsto (u, v) = (s(X - \varphi), sY)$$
$$\alpha = 3\varphi s, \qquad \beta = s$$
where $s = \sqrt{3\varphi^2 + A}^{-1}$

Table 3.7: Transformation from Weierstrass form into Montgomery form.

Montgomery	**Twisted Edwards**
$E_{M,\alpha,\beta} : \beta v^2 = u^3 + \alpha u + u$	$E_{E,a,d} : ax^2 + y^2 = 1 + dx^2y^2$
$\beta(\alpha^2 - 4) \neq 0$	$a \neq 0,\ d \neq 0,\ a \neq d$

$$(u, v) \longmapsto (x, y) = \left(\frac{u}{v}, \frac{u-1}{u+1}\right)$$
$$a = \frac{\alpha + 2}{\beta}, \qquad d = \frac{\alpha - 2}{\beta}$$

Table 3.8: Transformation from Montgomery form into Twisted Edwards form.

transformation in order to revert back to a Weierstrass curve that has previously been transformed into an Edwards curve using table 3.10. This is because in order to create the "Weierstrass into Edwards" transformation the term ax^2 in the Twisted Edwards curve equation, $ax^2+y^2 = 1+dx^2y^2$, must be substituted into $\hat{x}^2$. This makes $x = \dfrac{\hat{x}}{\sqrt{a}}$ and

$$ax^2 + y^2 = 1 + dx^2y^2 \quad \Longleftrightarrow \quad \hat{x}^2 + y^2 = 1 + d\left(\frac{\hat{x}}{\sqrt{a}}\right)^2 y^2.$$

So that $\hat{d} = \dfrac{d}{a}$ and $\hat{x}^2 + y^2 = 1 + \hat{d}\hat{x}^2y^2$. So, it is actually not the same variable and constant, "x" respectively "d", in table 3.6 and table 3.10. It is possible to revert back and forth between forms using the appropriate tables. Table 3.4

Weierstrass	**Twisted Edwards**
$E_{W,A,B} : Y^2 = X^3 + AX + B$	$E_{E,a,d} : ax^2 + y^2 = 1 + dx^2y^2$
$4A^3 + 27B^2 \neq 0$	$a \neq 0,\ d \neq 0,\ a \neq d$

1. $z^3 + Az + B = 0$ has one root $\varphi \in K$; and
2. $3\varphi^2 + A$ is a quadratic residue in K.

$$(X, Y) \longmapsto (x, y) = \left(\frac{(X - \varphi)}{Y}, \frac{s(X - \varphi) - 1}{s(X - \varphi)} \right)$$

$$a = \frac{3\varphi s + 2}{s}, \qquad d = \frac{3\varphi s - 2}{s}$$

where $s = \sqrt{3\varphi^2 + A}^{-1}$

Table 3.9: Transformation from Weierstrass form into Twisted Edwards form.

Weierstrass	**Edwards**
$E_{W,A,B} : Y^2 = X^3 + AX + B$	$E_{E,d} : x^2 + y^2 = 1 + dx^2y^2$
$4A^3 + 27B^2 \neq 0$	$d \neq \{0, 1\}$

1. $z^3 + Az + B = 0$ has one root $\varphi \in K$; and
2. $3\varphi^2 + A$ is a quadratic residue in K.

$$(X, Y) \longmapsto (x, y) = \left(\frac{\sqrt{\frac{3\varphi s + 2}{s}}(X - \varphi)}{Y}, \frac{s(X - \varphi) - 1}{s(X - \varphi) + 1} \right)$$

$$d = \frac{3\varphi s - 2}{3\varphi s + 2}$$

where $s = \sqrt{3\varphi^2 + A}^{-1}$

Table 3.10: Transformation from Weierstrass form into Edwards form.

and table 3.7 for the Weierstrass and Montgomery forms. Table 3.5 and table 3.8 for the Twisted Edwards and Montgomery forms. And table 3.6 and table 3.9 for the Weierstrass and Twisted Edwards forms. If we want to revert back to the Weierstrass form from the Edwards form after using the transformation

in table 3.10. The inverse transformation is given by

$$(x,y) \longmapsto (X,Y) = \left(\frac{s\varphi(1-y)+1+y}{s(1-y)}, \frac{\sqrt{\frac{3\varphi s+2}{2}\frac{1+y}{s(1-y)}}}{x} \right),$$

$$s = \frac{2(1+d)}{3\varphi(1-d)},$$

$$A = \left(\frac{3^2(1-d)^2}{2^2(1+d)^2} - 3 \right) \varphi^2 \quad \text{and} \quad B = \left(2 - \frac{3^2(1-d)^2}{2^2(1+d)^2} \right) \varphi^3.$$

To use this transformation we must know what φ is beforehand. For this reason we can only use this after we already used the transformation in table 3.10.

Now, let us finaly move on to the addition law for Edwards curve.

Definition 3.3.6 (Point addition on Edwards curve)**.** Let $E_{E,d} : x^2 + y^2 = 1 + dx^2y^2$ be an Edwards curve over the field K where $d \in K \setminus \{0,1\}$. The sum of the two points $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$ is

$$p_1 + p_2 = \left(\frac{x_1y_2 + y_1x_2}{1 + dx_1x_2y_1y_2}, \frac{y_1y_2 - x_1x_2}{1 - dx_1x_2y_1y_2} \right).$$

Remark. $d \in K \setminus \{0,1\}$ derives from the Twisted Edwards criteria. Since $a = 1$ in Edwards form, $d = 1$ would mean that $a = d$.

As hinted in the beginning of this section, the point addition on Edwards curves is the same regardless of the points being added. Compare it to the point addition on Weierstrass curves, 3.1.8, where there are several different cases depending on the points chosen. $(0,1)$ is the identity and the inverse of a point (x_1, y_1) is $(-x_1, y_1)$. There are some exceptional points that can occur when the denominators are 0. If d is not a square in the current field K then there will be no exceptional points.

Let us look at two simple examples of the point addition on a Edwards curve. The first over $\mathbb{R}$ (or more precise $\overline{\mathbb{R}}$) and the second over the finite field $\mathbb{F}_{11}$.

Example 3.3.7. Let $E_{E,-30} : x^2 + y^2 = 1 - 30x^2y^2$, $p_1 = (1,0)$ and $p_2 =$

$\left(\frac{1}{2}, \sqrt{\frac{3}{34}}\right)$. Then

$$p_3 = (1,0) + \left(\frac{1}{2}, \sqrt{\frac{3}{34}}\right) =$$

$$= \left(\frac{1\sqrt{\frac{3}{34}} + 0 \cdot \frac{1}{2}}{1 + (-30) \cdot 1 \cdot \frac{1}{2} \cdot 0 \cdot \sqrt{\frac{3}{34}}}, \frac{0\sqrt{\frac{3}{34}} - 1 \cdot \frac{1}{2}}{1 - (-30) \cdot 1 \cdot \frac{1}{2} \cdot 0 \cdot \sqrt{\frac{3}{34}}} \right)$$

$$= \left(\sqrt{\frac{3}{34}}, -\frac{1}{2}\right).$$

See figure 3.4.

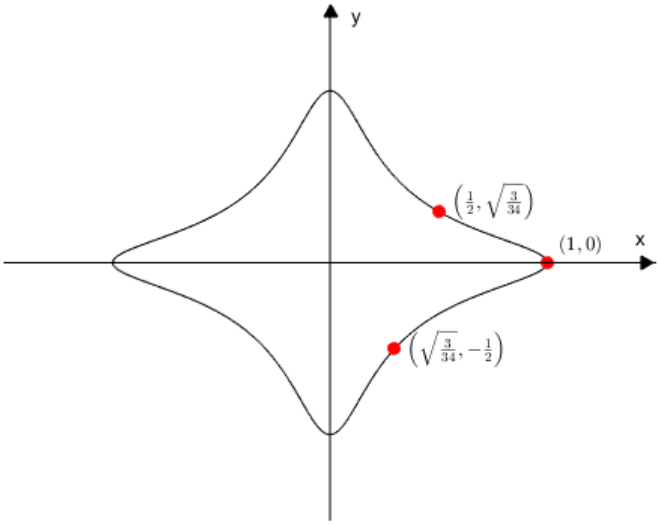

Figure 3.4: The curve $E_{E,-30} : x^2 + y^2 = 1 - 30x^2y^2$ and the points $p_1 = (1,0)$, $p_2 = \left(\frac{1}{2}, \sqrt{\frac{3}{34}}\right)$ and $p_3 = \left(\sqrt{\frac{3}{34}}, -\frac{1}{2}\right)$.

Example 3.3.8. Let $E_{E,7} : x^2 + y^2 = 1 + 7x^2y^2$ over the finite field $\mathbb{F}_{11}$. There are 16 points in $E_{E,7}(\mathbb{F}_{11})$,

$$(0,1),\ (0,10),\ (1,0),\ (2,4),\ (2,7),\ (3,3),\ (3,8),\ (4,2),\ (4,9),$$
$$(7,2),\ (7,9),\ (8,3),\ (8,8),\ (9,4),\ (9,7) \text{ and } (10,0).$$

Let $p_1 = (2,4)$ and let $p_2 = (7,2)$, then

$$\begin{aligned} p_3 = (2,4) + (7,2) &= \left(\frac{2 \cdot 2 + 4 \cdot 7}{1 + 7 \cdot 2 \cdot 7 \cdot 4 \cdot 2}, \frac{4 \cdot 2 - 2 \cdot 7}{1 - 7 \cdot 2 \cdot 7 \cdot 4 \cdot 2} \right) = \\ &= \left(\frac{32}{785}, \frac{-6}{-783} \right) = \left(\frac{10}{4}, \frac{5}{9} \right) = \\ &= (10 \cdot 3, 5 \cdot 5) = (8,3). \end{aligned}$$

Next, let us look at an example where we compare the results of the point addition on a Weierstrass curve and the point addition on its Edwards curve counterpart.

Example 3.3.9. Let $E_{W,2,-3} : Y^2 = X^3 + 2X - 3$, $P = (2,3)$ and $Q = (6,-15)$. We will be using transformation table 3.10. If K is $\mathbb{R}$ (or $\mathbb{C}$) then criteria 1 and 2 are satisfied. $z^3 + 2z - 3 = 0$ yield $z_1 = 1$ (and $z_{2,3} = -\frac{1}{2} \pm \frac{i\sqrt{11}}{2}$ if in $\mathbb{C}$). Let $\varphi = 1$. Then

$$s = \sqrt{3 \cdot 1^2 + 2}^{-1} = \frac{1}{\sqrt{5}} \quad \text{and} \quad d = \frac{3 \cdot 1 \cdot \frac{1}{\sqrt{5}} - 2}{3 \cdot 1 \cdot \frac{1}{\sqrt{5}} + 2} = \frac{12\sqrt{5} - 29}{11}$$

so that the Edwards curve is

$$E_{E, \frac{12\sqrt{5}-29}{11}} : x^2 + y^2 = 1 + \frac{12\sqrt{5} - 29}{11} x^2 y^2.$$

The points p and q are given by

$$P = (2,3) \longmapsto \left(\frac{\sqrt{\frac{3 \cdot 1 \cdot \frac{1}{\sqrt{5}} + 2}{\frac{1}{\sqrt{5}}}(2-1)}}{3}, \frac{\frac{1}{\sqrt{5}}(2-1) - 1}{\frac{1}{\sqrt{5}}(2-1) + 1} \right) =$$

$$= \left(\frac{\sqrt{3+2\sqrt{5}}}{3}, \frac{\sqrt{5}-3}{2} \right) = p$$

$$Q = (6,-15) \longmapsto \left(\frac{\sqrt{\frac{3 \cdot 1 \cdot \frac{1}{\sqrt{5}} + 2}{\frac{1}{\sqrt{5}}}(6-1)}}{-15}, \frac{\frac{1}{\sqrt{5}}(6-1) - 1}{\frac{1}{\sqrt{5}}(6-1) + 1} \right) =$$

$$= \left(-\frac{\sqrt{3+2\sqrt{5}}}{3}, \frac{3-\sqrt{5}}{2} \right) = q$$

Using the Elliptic Curve Addition Algorithm, 3.1.8, on $E_{W,2,-3}$ for the points P and Q yields

$$(2,3) + (6,-15) = \left(\frac{49}{4}, \frac{345}{8} \right)$$

so that $R = \left(\frac{49}{4}, \frac{345}{8} \right)$. Using the point addition on Edwards curve, 3.3.6, on $E_{E,\frac{12\sqrt{5}-29}{11}} : x^2 + y^2 = 1 + \frac{12\sqrt{5}-29}{11} x^2 y^2$ for the points p and q yield

$$p+q=$$

$$=\left(\frac{\left(\frac{\sqrt{3+2\sqrt{5}}}{3}\right)\left(\frac{3-\sqrt{5}}{2}\right)+\left(\frac{\sqrt{5}-3}{2}\right)\left(-\frac{\sqrt{3+2\sqrt{5}}}{3}\right)}{1+\left(\frac{12\sqrt{5}-29}{11}\right)\left(\frac{\sqrt{3+2\sqrt{5}}}{3}\right)\left(-\frac{\sqrt{3+2\sqrt{5}}}{3}\right)\left(\frac{\sqrt{5}-3}{2}\right)\left(\frac{3-\sqrt{5}}{2}\right)},\right.$$

$$\left.\frac{\left(\frac{\sqrt{5}-3}{2}\right)\left(\frac{3-\sqrt{5}}{2}\right)-\left(\frac{\sqrt{3+2\sqrt{5}}}{3}\right)\left(-\frac{\sqrt{3+2\sqrt{5}}}{3}\right)}{1-\left(\frac{12\sqrt{5}-29}{11}\right)\left(\frac{\sqrt{3+2\sqrt{5}}}{3}\right)\left(-\frac{\sqrt{3+2\sqrt{5}}}{3}\right)\left(\frac{\sqrt{5}-3}{2}\right)\left(\frac{3-\sqrt{5}}{2}\right)}\right)$$

$$=\left(\frac{6}{23}\sqrt{3+2\sqrt{5}},\frac{421-72\sqrt{5}}{389}\right)=r.$$

Using the transformation from Weierstrass to Edwards on R,

$$R=\left(\frac{49}{4},\frac{345}{8}\right)\longmapsto\left(\frac{\sqrt{\frac{3\cdot 1\cdot\frac{1}{\sqrt{5}}+2}{\frac{1}{\sqrt{5}}}}\left(\frac{49}{4}-1\right)}{\frac{345}{8}},\frac{\frac{1}{\sqrt{5}}\left(\frac{49}{4}-1\right)-1}{\frac{1}{\sqrt{5}}\left(\frac{49}{4}-1\right)+1}\right)=$$

$$=\left(\frac{6}{23}\sqrt{3+2\sqrt{5}},\frac{421-72\sqrt{5}}{389}\right)=r,$$

confirms that R and r represent the same point in their respective elliptic curve form and that the two addition algorithms are equivalent.

Chapter 4

Cryptography

The RSA cryptosystem was developed in the late 1970s and is based on prime numbers and factorisation of large integers. It was patented by its creators (R. Rivest, A. Shamir, L. Adleman) and was for a long time the only cryptosystem available for data transmission (It was actually developed earlier by people working for the British government, but the discoveries were classified well into the 90s.). In 1984, Hendrik Lenstra Jr. developed a factorisation method using elliptic curves. This was used (as are other factorisation methods) to break RSA and other cryptosystems but it also became the overall start for use of elliptic curves within cryptography. In 1985, Neal Koblitz and Victor Miller suggested, independently of each other, that elliptic curves could be used in cryptosystems and that elliptic curves might be able to utilise smaller keys than other cryptosystems but with the same level of security. For a long time Elliptic Curve Cryptography (ECC) had no patents associated with it, this had both pros and cons. It meant that more people were able to develop cryptosystems based on elliptic curves but these cryptosystems were not (always) as thoroughly tested and scrutinised as e.g. RSA and therefore it was questioned whether they were secure or not. For example, there was a system based on so called "supersingular elliptic curves". These curves were chosen because they seemed efficient but later it turned out that they were vulnerable to certain attacks. Examples like this gave ECC a bad reputation for a long time even though other cryptosystems also have special cases that need to be avoided in order for them to be considered safe. However, today ECC is held in much higher regard and is viewed to have the same level of security as RSA but with the smaller keys (as predicted by Koblitz and Miller). It is also used as part of the basis for many cryptocurrencies, where curves of Weierstrass, Montgomery and Edwards form are used. For example the most famous cryptocurrency, Bitcoin, uses the

Weierstrass curve $E : y^2 = x^3 + 7$.

4.1 Elliptic Curve Algorithms

In this section we will take a short look at how elliptic curves are used in practice. We will look at the underlying problem that constitutes the basis for all cryptological applications and a short example on how to speed up the calculations related to it before describing some of the cryptosystems that utilises elliptic curves. The focus will not be on the differences or origin of different cryptosystems, nor on how adversaries can try to break them. For an overview of these topics [3] covers all these things. Indeed, [3] is the source for the content of this section.

First let us look at the problem that is the basis for all ECC.

Definition 4.1.1. Let E be an elliptic curve over the finite field $\mathbb{F}_p$ and let P and Q be points on $E(\mathbb{F}_p)$. The *Elliptic Curve Discrete Logarithm Problem* (ECDLP) is the problem of finding an integer n such that $Q = nP$. We denote this integer n by

$$n = \log_P(Q)$$

and we call n the *elliptic discrete logarithm of Q with respect to P.*

Remark. It is not guaranteed that there exists an n such that $Q = nP$ for arbitrary points P and Q in $E(\mathbb{F}_p)$. But for different cryptography applications Q is calculated from nP so in practice $Q = nP$ will be defined.

For many types of cryptosystems, the process of encrypting and decrypting the data or "the secret message" relies on keys created by some public number or, in the case of ECC, point and one or several secret numbers. This is where the ECDLP appears. But in order for the ECDLP not to be guessed or easily calculated by an adversary n needs to be large. Since

$$nP = \underbrace{P + P + \ldots + P}_{\text{n times}},$$

calculating nP by first calculating $2P, 3P, 4P, \ldots, (n-1)P$ is very inefficient for large values of n. There are several methods and algorithms that reduce the amount of calculations needed in order to compute nP. We will limit ourselves to just one where the idea is to only add some particular doublings of the chosen point.

Algorithm 4.1.2 (Double and add algorithm)**.** Let $P \in E(\mathbb{F}_p)$ and $n \geq 1$.

1. Set $Q = P$ and $R = \mathcal{O}$.

2. Loop while $n > 0$.

    3. If $n \equiv 1 \pmod 2$, set $R = R + Q$.
    4. Set $Q = 2Q$ and $n = \lfloor n/2 \rfloor$.
    5. If $n > 0$, continue with loop at step 2.

6. Return the point R, which equals nP.

Remark. Here $\mathcal{O}$ represents the identity. This means that for a curve on Weierstrass form $\mathcal{O}$ is the point at infinity but for a curve in Edwards form it is the point $(0, 1)$.

Let us look at an example for a Weierstrass curve.

Example 4.1.3. Let $E : X^3 + 2X - 3$ in $\mathbb{F}_{23}$, $P = (16, 10)$ and $n = 21$. What is $21 \cdot (16, 10)$? Let us look at the first two iterations of the double and add algorithm. The first iteration:

Starting values: $n = 21, \quad Q = P = (16, 10), \quad R = \mathcal{O}$
New values: $n = 21 \equiv 1 \pmod 2, \qquad R := R + Q = \mathcal{O} + (16, 10) = (16, 10)$
$Q := 2Q = (16, 10) + (16, 10) = (7, 3), \qquad n := \lfloor n/2 \rfloor = \lfloor 21/2 \rfloor = 10$

The second iteration:

Starting values: $n = 10, \quad Q = (7, 3), \quad R = (16, 10)$
New values: $n = 10 \not\equiv 1 \pmod 2, \qquad R = (16, 10) \quad$ (No change.)
$Q := 2Q = (7, 3) + (7, 3) = (13, 9), \qquad n := \lfloor n/2 \rfloor = \lfloor 10/2 \rfloor = 5$

The remaining iterations follow the same procedure and will be omitted. The completed double and add algorithm is compiled in table 4.1.

Iteration i	n	$Q = 2^i P$	R
0	$n = 21$	$Q = P = (16, 10)$	$R = \mathcal{O}$
1	10	$(7, 3)$	$(16, 10)$
2	5	$(13, 9)$	$(16, 10)$
3	2	$(13, 14)$	$(12, 22)$
4	1	$(13, 9)$	$(12, 22)$
5	0	$(13, 14)$	$(6, 15)$

Table 4.1: Double and add algorithm for the point $P = (16, 10)$ on $E : X^3 + 2X - 3$ in $\mathbb{F}_{23}$.

We conclude that $21 \cdot (16, 10) = (6, 5)$.

It took 5 iterations of the double and add algorithm to solve our example. That includes 5 additions of $Q + Q$ and 3 additions of $R + Q$ as well as some calculations regarding the value of n. Compare that to the 20 point additions it would have required to calculate $2P, 3P, \ldots, 20P, 21P$. The difference becomes even more apparent the greater the value of n. Also, remember that point addition on a Weierstrass curve requires several steps in itself depending on the points added, see 3.1.8. Using an Edwards curve will, in general, require fewer "hidden" steps but it does not change the number of iterations needed in the double and add algorithm.

There are algorithms that are designed specifically for reducing the number of calculations needed for specific types of curves. For example [8], [9] and [11] presents several ways to reduce calculations on Edwards and Twisted Edwards curves but these will not be explored in this essay. Now let us move on to some of the ECC applications.

All algorithms presented here have corresponding versions made for using prime numbers rather than points so the names and processes might be familiar if the reader has seen the "prime number" counterparts previously.

First up is the *Elliptic Diffie-Hellman Key Exchange* presented in table 4.2. The purpose of this algorithm is that the participants, here called Alice and Bob, can agree on a secret key using an unsecure channel. The key can later be used to, for example, encrypt messages using some sort of symmetric cipher. "Symmetric" means that the same key is used both for encryption and decryption.

Public parameter creation	
A trusted party chooses and publishes a (large) prime p, an elliptic curve E over $\mathbb{F}_p$, and a point P in $E(\mathbb{F}_p)$.	
Private computations	
Alice	**Bob**
Chooses a secret integer n_A. Computes the point $Q_A = n_A P$.	Chooses a secret integer n_B. Computes the point $Q_B = n_B P$.
Public exchange of values	
Alice sends Q_A to Bob $\longrightarrow Q_A$	
$Q_B \longleftarrow$ Bob sends Q_B to Alice	
Further private computations	
Alice	**Bob**
Computes the point $n_A Q_B$.	Computes the point $n_B Q_A$.
The shared secret value is $n_A Q_B = n_A(n_B P) = n_B(n_A P) = n_B Q_A$.	

Table 4.2: Diffe-Hellman key exchange.

Using table 4.2, let us look at a simple example using a Weierstrass curve.

Example 4.1.4. Let $p = 23$, $E(\mathbb{F}_{23}) : Y^2 = X^3 + 2X - 3$ and $P = (16, 10) \in E(\mathbb{F}_{23})$ be the public parameters. Alice chooses her secret integer $n_A = 19$ and computes

$$Q_A = n_A P = 19 \cdot (16, 10) = (12, 1).$$

Simultaneously, Bob chooses his secret integer $n_B = 17$ and computes

$$Q_B = n_B P = 17 \cdot (16, 10) = (12, 22).$$

Alice sends Q_A to Bob and Bob sends Q_B to Alice. Alice computes $n_A Q_B = 19 \cdot (12, 22)$ and Bob computes $n_B Q_A = 17 \cdot (12, 1)$ so that their shared secret value is

$$n_A Q_B = 19 \cdot (12, 22) = n_B Q_A = 17 \cdot (12, 1) = (16, 13).$$

Next is the *Elliptic Elgamal Public Key Cryptosystem*, presented in table 4.3. The Elgamal cryptosystem is a way to encrypt and decrypt messages, rather than just agreeing on a key as with Diffie-Hellman.

Public parameter creation	
A trusted party chooses and publishes a (large) prime p, an elliptic curve E over $\mathbb{F}_p$, and a point P in $E(\mathbb{F}_p)$.	
Alice	**Bob**
Key creation	
Choose a private key n_A. Compute $Q_A = n_A P$ in $E(\mathbb{F}_p)$. Publish the public key Q_A.	
Encryption	
	Choose plaintext $M \in E(\mathbb{F}_p)$. Choose a random element k. Use Alice's public key Q_A to compute $C_1 = kP \in E(\mathbb{F}_p)$, and $C_2 = M + kQ_A \in E(\mathbb{F}_p)$. Send ciphertext (C_1, C_2) to Alice.
Decryption	
Compute $C_2 - n_A C_1 \in E(\mathbb{F}_p)$. This quantity is equal to M.	

Table 4.3: Elgamal cryptosystem.

Let us also look at a simple example of the Elgamal cryptosystem, this time using an Edwards curve.

Example 4.1.5. Let $p = 37$, $E(\mathbb{F}_{37}) : x^2 + y^2 = 1 + 5x^2y^2$ and $P = (4, 15) \in E(\mathbb{F}_{37})$ be the public parameters. Alice chooses a private key $n_A = 27$ and computes

$$Q_A = n_A P = 27 \cdot (4, 15) = (32, 6).$$

Alice publishes the public key $Q_A = (32, 6)$. Bob chooses the plaintext $M = (19, 16) \in E(\mathbb{F}_{37})$ and the random element $k = 15$. Then, using Alice public key he computes

$$C_1 = kP = 15 \cdot (4, 15) = (4, 22)$$
$$C_2 = M + kQ_A = (19, 16) + 15(32, 6) = (19, 16) + (32, 31) = (6, 5).$$

Bob sends the ciphertext $(C_1, C_2) = ((4, 22), (6, 5))$ to Alice. Alice computes

$$\begin{aligned} C_2 - n_A C_1 &= (6, 5) - 27 \cdot (4, 22) = \\ &= (6, 5) - (32, 31) = (6, 5) + (5, 31) = (19, 16). \end{aligned}$$

And thus Alice can read Bobs secret message $M = (19, 16)$.

Remark. Recall that the negative of a point on an Edwards curve is $-(x, y) = (-x, y)$. This is why $-(32, 31) = (-32, 31) = (5, 31)$ in our example above.

The *Elliptic Curve Digital Signature Algorithm*, ECDSA, is a way to sign, for example, a digital document so that anyone viewing the signature can confirm that the author are who they claim to be and that the document is not a forgery. The algorithm is presented in table 4.4. We call the signer Samantha and the verifier Victor.

Remark. The order of a point, referred to in table 4.4, is the smallest number of times a point needs to be added to itself in order to reach the identity. I.e. the smallest n that solves $nP = \mathcal{O}$, where $\mathcal{O}$ is the identity. Compare it to number theory where the smallest integer x that solves $a^x \equiv 1 \pmod{n}$, where a and n are relatively prime, is called the order of a modulo n.

Let us look at a simple example of ECDSA using an Edwards curve.

Example 4.1.6. Let $E(\mathbb{F}_{43}) : x^2 + y^2 = 1 + 5x^2y^2$ and $G = (7, 12)$ with prime order $q = 13$ be the public parameters. Samantha chooses her secret signing key $s = 11$ (where $1 < s < q - 1 = 12$) and computes

$$V = sG = 11 \cdot (7, 12) = (35, 38)$$

and publish V. Samantha choose document $d = 8 \mod q = 13$ and the random element $e = 9 \mod q = 13$. She then computes

$$eG = 9 \cdot (7, 12) = (41, 39)$$

Public parameter creation	
A trusted party chooses a finite field $\mathbb{F}_p$, an elliptic curve $E(\mathbb{F}_p)$, and a point $G \in E(\mathbb{F}_p)$ of large prime order q.	
Samantha	**Victor**
Key creation	
Choose a secret signing key $1 < s < q-1$. Compute $V = sG \in E(\mathbb{F}_p)$. Publish the verification key V.	
Signing	
Choose document $d \pmod q$. Choose random element $e \pmod q$. Compute $eG \in E(\mathbb{F}_p)$ and then, $s_1 = x(eG) \pmod q$ and $s_2 \equiv (d + ss_1)e^{-1} \pmod q$. Publish the signature (s_1, s_2)	
Verification	
	Compute $v_1 \equiv ds_2^{-1} \pmod q$ and $v_2 \equiv s_1 s_2^{-1} \pmod q$. Compute $v_1 G + v_2 V \in E(\mathbb{F}_p)$ and verify that $x(v_1 G + v_2 V) \pmod q = s_1$.

Table 4.4: Elliptic Curve Digital Signature Algorithm (ECDSA).

and let s_1 be the x-coordinate of $eG \mod q$, that is

$$s_1 = 41 = 2 \mod 13,$$

and let $s_2 \equiv (d + ss_1)e^{-1} \pmod q$, where e^{-1} is the inverse of $e \pmod q$, that is

$$s_2 \equiv (8 + 11 \cdot 2) \cdot 3 \equiv 90 \equiv 12 \pmod{13}.$$

Samantha's final step is to publish $d = 8$ and her signature $(s_1, s_2) = (2, 12)$. In order to verify that the signature is valid Victor computes $v_1 \equiv ds_2^{-1} \pmod q$ and $v_2 \equiv s_1 s_2^{-1} \pmod q$. That is

$$v_1 \equiv 8 \cdot 12 \equiv 96 \equiv 5 \pmod{13}$$
$$v_2 \equiv 2 \cdot 12 \equiv 24 \equiv 11 \pmod{13}.$$

Finally, Victor computes

$$v_1 G + v_2 V = 5 \cdot (7, 12) + 11 \cdot (35, 38) = (18, 28) + (2, 39) = (41, 39)$$

and verifies that the resulting x-coordinate (mod q) is equal to s_1, that is

$$x(41, 39) = 41 \equiv 2 \pmod{13} = s_1.$$

To put into perspective how large keys are in real applications of different ECCs (and to highlight how far from secure the examples in this chapter are) here are the public parameters used in the ECDSA for the cryptocurrency Bitcoin. Bitcoin uses the Weierstrass curve

$$E : Y^2 = X^3 + 7$$

in the finite field $\mathbb{F}_p$, where

$$\begin{aligned} p &= 2^{256} - 2^{32} - 2^9 - 2^8 - 2^7 - 2^6 - 2^4 - 1 = \\ &= 115792089237316195423570985008687907853269984665640564039457584007908834671663. \end{aligned}$$

The public point $G = (x, y) \in E(\mathbb{F}_p)$ where the x- and y-coordinates are

$$x = 55066263022277343669578718895168534326250603453777594175500187360389116729240$$

$$y = 32670510020758816978083085130507043184471273380659243275938904335757337482424.$$

4.2 Practical Uses and the Future

Bitcoin brings us nicely to the topic of *blockchains*, which is the technology behind Bitcoin and other cryptocurrencies as well as a variety of other applications, which utilises ECDSA as part of its structure. The sources for this section are [3] and [12].

A blockchain is a form of decentralised database which has a tree structure and the data stored in it is immutable. Each node, called *block*, in the tree contains the address to its parent node so that all blocks form a chain from the outermost blocks all the way back to the first block, called the *genesis block*. Since each block has the ability to retrace the whole chain, the chain holds the history of every action ever performed on the chain. The idea is that the data is stored with all participants, called nodes (not to be confused with the block nodes), and in order to make changes to the chain, i.e. adding blocks, a majority

of the nodes must approve or verify the change. The participants are the people or computers that utilise the blockchain. In the case of cryptocurrencies the participants are those who own and trade with the currency. In practice the tree structure consists of a long chain with some short branches. The short branches are blocks that got added to the chain but never achieved majority approval and therefore got discarded. The simplest general block structure consists of three things:

- the address of the previous block in the form of a hash,
- the payload, i.e. the data specific to the purpose of the particular blockchain, and
- a signature belonging to the creator of the block.

It is in the signing and verifying step where ECDSA is used. Each participant owns a pair of keys, a public and a private. When a participant creates a new block they sign it with their private key and then the block is broadcast to the network. Nodes that receive the new block verify its authenticity using the public key and subsequently add it to the blockchain. When a majority of nodes have approved of the new block it is considered valid and becomes a permanent new part of the chain.

This technology has a variety of uses beyond cryptocurrency. Due to the nature of its immutability it is also suitable for many applications within finance such as smart contracts, taxation, international trade etc. which the reader can find more information about in [13].

It is worth mentioning that there are concerns regarding ethics and environmental impact for some of the blockchain applications. For example the energy consumption for the Bitcoin blockchain was estimated to approximately 70 TWh per year in 2018. As such there are ongoing attempts to change certain parts of different blockchain algorithms to make them more energy efficient.

But in summary it is fair to say that blockchain technology is a highly relevant and current topic that can be expected to keep evolving and finding new areas of applications.

And what of ECC? The greatest "threat" to ECC, as well as other form of cryptography, is the development of quantum computers. The computational power of quantum computers are estimated to vastly surpass the power of today's computers. Thus, the cryptosystems of today would go from hard to easy to solve and become obsolete. But until then ECC provides the high security with shorter keys that Koblitz and Miller predicted.

Bibliography

[1] D. A. Cox, J. B. Little, D. O'Shea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*, third edition, Springer, 2007

[2] D. S. Dummit, R. M. Foote, *Abstract algebra*, Englewood Cliffs, N.J. Prentice Hall, 1991

[3] J. Hoffstein, J. C. Pipher, J. H. Silverman, *An Introduction to Mathematical Cryptography*, Springer, New York, London, 2008

[4] G. A. Jones, D. Singerman, *Complex functions: an algebraic and geometric viewpoint*, Cambridge Univ. Press, Cambridge, 1987

[5] M. Reid, *Undergraduate algebraic geometry*, Cambridge Univ. Press, Cambridge, 1988

[6] R. J. Walker, *Algebraic Curves*, Springer-Verlag, New York, 1978

[7] H. M. Edwards, *A Normal Form for Elliptic Curves*, Bulletin of the American Mathematical Society 44, 393–422. URL: `https://www.ams.org/journals/bull/2007-44-03/S0273-0979-07-01153-6/S0273-0979-07-01153-6.pdf`, 2007

[8] D. J. Bernstein, T. Lange *Faster Addition and Doubling on Elliptic Curves*, In: Kurosawa, K. (eds) Advances in Cryptology – ASIACRYPT 2007. ASIACRYPT 2007. Lecture Notes in Computer Science, vol 4833. Springer, Berlin, Heidelberg. URL: `https://doi.org/10.1007/978-3-540-76900-2_3`, 2007

[9] D. J. Bernstein, P. Birkner, M. Joye, T. Lange, C. Peters, *Twisted Edwards Curves*, In: Vaudenay, S. (eds) Progress in Cryptology – AFRICACRYPT 2008. AFRICACRYPT 2008. Lecture Notes in Computer Science, vol

5023. Springer, Berlin, Heidelberg. URL: https://doi.org/10.1007/978-3-540-68164-9_26, 2008

[10] K. Okeya, H. Kurumatani, K. Sakurai, *Elliptic Curves with the Montgomery-Form and Their Cryptographic Applications*, In: Imai, H., Zheng, Y. (eds) Public Key Cryptography. PKC 2000. Lecture Notes in Computer Science, vol 1751. Springer, Berlin, Heidelberg. URL: https://doi.org/10.1007/978-3-540-46588-1_17, 2000

[11] H. Hisil, K. KH. Wong, G. Carter, E. Dawson, *Twisted Edwards Curves Revisited*, In: Pieprzyk, J. (eds) Advances in Cryptology - ASIACRYPT 2008. ASIACRYPT 2008. Lecture Notes in Computer Science, vol 5350. Springer, Berlin, Heidelberg. URL: https://doi.org/10.1007/978-3-540-89255-7_20, 2008

[12] F. Knirsch, A. Unterweger, D. Engel, *Implementing a blockchain from scratch: why, how, and what we learned.*, EURASIP J. on Info. Security 2019, 2 (2019). URL: https://doi.org/10.1186/s13635-019-0085-3, 2019

[13] D. Kimani, K. Adams, R. Attah-Boakye S. Ullah J. Frecknall-Hughes, J. Kim, *Blockchain, business and the fourth industrial revolution: Whence, whither, wherefore and how?*, In: Technological Forecasting and Social Change, vol 161. URL: https://doi.org/10.1016/j.techfore.2020.120254, 2020

Appendix A

Codes

Presented in this appendix are codes that were produced in order to help create some of the examples in this essay. The codes are written in MATLAB.

Code 1: Determines whether or not a point is on an Edwards curve and returns true or false.

```
% Determines whether or not a point is on an Edwards
    curve
function [output]=onEdwardsCurve(point,d)
if point(1)^2+point(2)^2==1+d*point(1)^2*point(2)^2
    output=true;
else
    output=false;
end
end
```

Code 2: Performs the point addition for an Edwards curve and returns the resulting point. Throws an error if the curve is invalid.

```
% Performs the point addition on an Edwards curve
function [x3,y3]=edwardsAddition(point1,point2,d)
if d^5==d
    error('Invalid curve. d may not be equal to 0, 1
        or -1')
else
    x3=(point1(1)*point2(2)+point1(2)*point2(1))/(1+d*
        point1(1)*point1(2)*point2(1)*point2(2));
```

```
    y3=(point1(2)*point2(2)-point1(1)*point2(1))/(1-d*
        point1(1)*point1(2)*point2(1)*point2(2));
end
end
```

Code 3: Point addition for an Edwards curve. Returns the resulting point. Throws an error if either of the points are not on the curve.

```
%% Point addition on Edwards curve
function edwardsCurveAddition(point1,point2,d)
if onEdwardsCurve(point1,d)==false
    error(sprintf('The point (%d, %d) is not on the
        specified curve.', point1))
elseif onEdwardsCurve(point2,d)==false
    error(sprintf('The point (%d, %d) is not on the
        specified curve.', point2))
else
    [x3,y3]=edwardsAddition(point1,point2,d);
end
result=[x3,y3]
end
```

Code 4: Calculates and returns a modular inverse. Throws an error if the numbers are not relatively prime.

```
%% Modular inverse
%Calculates an inverse for a modulo m
function [inverse] = modInverse(a,m)
    if gcd(a,m) ~= 1
        error('No solution. Your chosen values are not
            relatively prime.')
    else
        x=1;
        while mod(a*x,m) ~= 1
            x = x+1;
        end
        inverse = x;
    end
end
```

Code 5: Determines whether or not a point is on an Edwards curve over a finite field and returns true or false.

```
% Determines whether or not a point is on a given
% Edwards curve over a finite field
function [output] = on_Finite_E_Curve(point, d,
    finite_Field)
curveValue = point(1)^2+point(2)^2-1-d*point(1)^2*
    point(2)^2;
if mod(curveValue, finite_Field)==0
     output = true;
else
     output = false;
end
end
```

Code 6: Performs the addition algorithm for an Edwards curve over a finite field and returns the resulting point.

```
% Performs the point addition for an Edwards curve
    over a finite field
function [x3,y3]=finite_E_Add(P, Q, d, finite_Field)

if mod(1+d*P(1)*Q(1)*P(2)*Q(2),finite_Field)==0 || mod
    (1-d*P(1)*Q(1)*P(2)*Q(2),finite_Field)==0
     error('One of the divisors are 0.')
else
denominator_xcoord = modInverse(1+d*P(1)*Q(1)*P(2)*Q
    (2), finite_Field);
denominator_ycoord = modInverse(1-d*P(1)*Q(1)*P(2)*Q
    (2), finite_Field);

x3 = mod((P(1)*Q(2)+P(2)*Q(1))*denominator_xcoord,
    finite_Field);
y3 = mod((P(2)*Q(2)-P(1)*Q(1))*denominator_ycoord,
    finite_Field);
end
```

Code 7: Point addition for an Edwards curve over a finite field. Displays the resulting point. Throws errors if any of the inputs are invalid.

```
%% Point addition for an Edwards curve over a finite
    field
function finite_E_Addition(point_1, point_2, d,
    finite_Field)
```

```
if mod(d,finite_Field)==0 || mod(d,finite_Field)==1
    error('Not a valid Edwards Curve. d may not be
        equal to 0 or 1 in your chosen field.')
elseif mod(d.^((finite_Field-1)/2),finite_Field)==1
    error('Not a vaild Edwards Curve. d may not be a
        quadratic residue in your chosen field.')
elseif isprime(finite_Field) == false
    error('Not a finite field.')
elseif on_Finite_E_Curve(point_1, d, finite_Field) ==
   false
    error(sprintf('The point (%d, %d) is not on the
        chosen curve.', point_1))
elseif on_Finite_E_Curve(point_2, d, finite_Field) ==
   false
    error(sprintf('The point (%d, %d) is not on the
        chosen curve.', point_2))
else [x3,y3]=finite_E_Add(point_1, point_2, d,
   finite_Field);
end
fprintf('(%d, %d) + (%d, %d) is equal to (%d, %d) \n',
    point_1,point_2,[x3,y3])
end
```

Code 8: The double and add algorithm for an Edwards curve over a finite field. Displays the resulting point. Throws errors if any of the inputs are invalid.

```
%% Double and Add algorithm for an Edwards curve
function doubleAndAdd_Edwards(scalar, point, d,
   finite_Field)
if isprime(finite_Field) == false
    error('Not a finite field.')
elseif mod(d,finite_Field)==0 || mod(d,finite_Field)
   ==1
    error('Not a valid Edwards Curve. d may not be
        equal to 0 or 1 in your chosen field.')
elseif mod(d.^((finite_Field-1)/2),finite_Field)==1
    error('d is a quadratic residue in your chosen
        field, therefore not all additions are defined.
        ')
elseif on_Finite_E_Curve(point, d, finite_Field) ==
   false
```

```
    error(sprintf('The point (%d, %d) is not on the
       chosen curve.', point))
else
    Q=point;
    R=[0 1];
    n=scalar;
    while n>0
        if mod(n,2) == 1
            [R(1),R(2)]=finite_E_Add(R,Q,d,
               finite_Field);
        end
        [Q(1),Q(2)]=finite_E_Add(Q,Q,d,finite_Field);
        n=floor(n/2);
    end
end
fprintf('%d times (%d, %d) is (%d, %d) \n', scalar,
   point,R)
end
```