Radical isogenies and modular curves

Pribanić, Valentina

University of Zagreb

FACULTY OF SCIENCE

DEPARTMENT OF MATHEMATICS

Valentina Pribanić

Radical isogenies and modular curves

DOCTORAL DISSERTATION

Zagreb, 2024.

University of Zagreb

FACULTY OF SCIENCE

DEPARTMENT OF MATHEMATICS

Valentina Pribanić

Radical isogenies and modular curves

DOCTORAL DISSERTATION

prof. dr. sc. Matija Kazalicki

Zagreb, 2024.

Sveučilište u Zagrebu

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Valentina Pribanić

Radikalne izogenije i modularne krivulje

DOKTORSKI RAD

prof. dr. sc. Matija Kazalicki

Zagreb, 2024.

ZAHVALA

Želim se zahvaliti svom mentoru, profesoru Matiji Kazalickom koji je u ovih nekoliko godina zajedničkog rada pokazao veliku razinu strpljenja i želje, na početku u pronalasku područja rada zanimljivog za nas oboje, a kasnije u brojnim iteracijama pisanja ovog rada. Pokazao je i puno razumijevanja za razna pitanja koja su ponekad, usudim se sada reći, bila i ispod razine doktorskog studenta. Naučila sam puno od njega o znanstvenom istraživanju i prevođenju raspisanog u strogi matematički tekst.

Želim se zahvaliti i profesoru Andreju Dujelli, koji je (ne)posredno bio dio mog puta od samog početka, prvo kroz diplomski rad, čija tema me je usmjerila prema radu na području kriptografije, zatim kod prijedloga mentora, a pri kraju i čitanjem tematike mog rada kada je još bio u obliku članka.

Hvala mojoj obitelji na njihovoj podršci tijekom studija i razumijevanju za sve propuštene vikende i radne akcije.

Hvala mojim kolegama na poslu (iz Odjela i šire) koji su tijekom ovih godina iskazali veliku količinu razumijevanja i susretljvosti kad god je ona bila potrebna. Posebno hvala Ani.

Hvala Nikoli i Teu jer su znali kada ne treba pitati kako ide. Hvala Ivani i Ivani na savjetima iz vlastitih iskustava.

SUMMARY

This thesis explores the connection between radical isogenies and modular curves. Radical isogenies are formulas designed to calculate chains of isogenies of fixed small degree N, introduced by Castryck, Decru, and Vercauteren in [17]. One significant advantage of radical isogeny formulas over other formulas with a similar purpose is that they eliminate the need to generate a point of order N that generates the kernel of the isogeny. While the authors originally developed radical isogeny formulas using elliptic curves in Tate normal form, in [54] Onuki and Moriya proposed radical isogeny formulas of degrees 3 and 4 on Montgomery curves and attempted to obtain a less complex form of radical isogenies using enhanced elliptic and modular curves.

In this thesis, we will first translate the original setup of radical isogenies in Tate normal form into the language of modular curves. Second, we will solve an open problem introduced by Onuki and Moriya regarding radical isogeny formulas on $X_0(N)$.

Chapter 1 provides the necessary background, definitions, and results regarding group theory, elliptic curves, modular curves, etc.

Chapter 2 gives a short introduction to cryptography and an overview of post-quantum cryptography, mainly focusing on the isogeny-based post-quantum cryptography.

Chapter 3 introduces radical isogenies and radical isogenies on Montgomery curves.

The last two chapters are based on the author's paper [58]. In Chapter 4, we generalize radical isogenies using modular curves. In the final Chapter 5, we extend the setting from the previous Chapter 4 to include the modular curve $X_0(N)$ and discuss the existence of radical isogeny formulas in this extended setting.

Keywords: elliptic curves, isogenies, modular curves, enhanced elliptic curves, isogeny-based post-quantum cryptography, radical isogenies.

SAŽETAK

Ovaj rad istražuje vezu između radikalnih izogenija i modularnih krivulja. Radikalne izogenije su formule namijenjene izračunu lanaca izogenija, gdje je svaka izogenija u tom lancu malog, fiksnog stupnja N. Ove formule se u literaturi prvi put pojavljuju 2020. godine u članku [17], autora Castrycka, Decrua i Vercauterena. Ono što razlikuje radikalne izogenije od sličnih formula iste namjene, je to da kod njih ne postoji potreba za poznavanjem ili generiranjem točke reda N na eliptičkoj krivulji koja generira jezgru izogenije. Radikalne izogenije su originalno razvijene za krivulje u Tateovoj normalnoj formi, a kasnije su Onuki i Moriya, u svom članku [54], pokazali da postoje i radikalne izogenije na Mongomeryjevim krivuljama stupnja 3 i 4, te su pokušali definirati jednostavniji oblik radikalnih izogenija koristeći teoriju modularnih i obogaćenih eliptičkih krivulja.

Dva su glavna rezultata ovog rada. Prvo, pokazat ćemo da se radikalne izogenije u Tateovoj normalnoj formi mogu generalizirati koristeći jezik modularnih krivulja. Drugo, riješit ćemo otvoreni problem, koji se prvi put u literaturi pojavio u [54], a tiče se radikalnih izogenija na modularnoj krivulji $X_0(N)$.

U poglavlju 1 definiramo osnovne pojmove i rezultate potrebne za ostatak rada, a direktno vezane za teoriju grupa, eliptičke krivulje, modularne krivulje itd.

U poglavlju 2 dan je pregled osnovnih pojmova povezanih s kriptografijom, gdje se posebno fokusiramo na post-kvantnu kriptografiju i post-kvantnu kriptografiju temeljenu na izogenijama.

U poglavlju 3 definiramo radikalne izogenije i radikalne izogenije na Montgomeryjevim krivuljama.

U zadnja dva poglavlja, koja prate autoričin članak [58], iskazujemo glavne rezultate ovog rada. U poglavlju 4 generaliziramo pojam radikalnih izogenija koristeći teoriju modularnih krivulja. Zatim, u zadnjem poglavlju 5, dodatno proširujemo generalizaciju

iz poglavlja 4 kako bismo uključili i modularnu krivulju $X_0(N)$, te razmatramo mogućnost postojanja radikalnih izogenija u tom proširenom kontekstu.

Ključne riječi: eliptičke krivulje, izogenije, modularne krivulje, obogaćene eliptičke krivulje, post-kvantna kriptografija bazirana na izogenijama, radikalne izogenije.

CONTENTS

INTRODUCTION

Post-quantum cryptography (PQC) is an area of cryptography focused on developing cryptosystems resistant to attacks from both classical and quantum computers. These systems rely on hard mathematical problems different from the integer factorization problem or (elliptic-curve) discrete logarithm problem, which most current cryptographic algorithms rely on. PQC includes various approaches to cryptography, such as lattice, code, multivariate, hash, and isogeny-based cryptography.

The first isogeny-based cryptosystem was proposed by Couveignes in 1997 [25], and it was independently rediscovered by Rostovtsev and Stolbunov in 2006 [61]. They described a non-interactive key exchange using ordinary elliptic curves. This scheme is commonly referred to as CRS. This field gained new momentum in 2011 with the proposal of SIDH by De Feo and Jao in [41], the supersingular isogeny Diffie-Hellman key exchange. A variant of this algorithm called SIKE was a promising candidate for NIST[1] PQC standardization, but it was broken in several independent papers in August 2022 [15, 47, 60]. In 2018, Castryck, Lange, Martindale, Panny, and Renes introduced CSIDH [18], a non-interactive key exchange protocol that adapts the CRS protocol to supersingular elliptic curves. The previously mentioned attacks on SIDH cannot be applied to CSIDH.

Compared to other post-quantum protocols, smaller key and ciphertext sizes are the main advantages of isogeny-based cryptography. On the other hand, the main disadvantage of isogeny-based protocols has been the high computational cost of encryption and decryption. These advantages and disadvantages are particularly evident in digital signatures. SQISign, introduced in 2020 [28], is among the most promising and compact isogeny-based digital signatures. It has seen some speed improvements in 2022 [29], but

[1] Short for National Institute of Standards and Technology.

despite this, it is still several orders of magnitude slower than other post-quantum signature schemes.

Protocols like CRS, CSIDH, or, for example, Charles, Goren, and Lauter's hash function [20], share the need to compute isogenies of low degree in a finite field. An isogeny can be computed from the coordinates of the points in its kernel using Vélu's formulas [67]. To improve and accelerate isogeny computation, various approaches and variants of Vélu's formulas have been proposed for different curve models, such as Montgomery curves in [23], Edwards curves [19,43], and Hessian curves [57]. An algorithm by Bernstein, De Feo, Leroux, and Smith [12] reduces the cost of computation of isogeny of degree N from $\mathcal{O}(N)$ to $\widetilde{\mathcal{O}}(\sqrt{N})$ and can be applied to Huff's curves [71].

Radical isogenies are formulas designed for computing a chain of isogenies of the same small degree between elliptic curves over a finite field, first introduced by Castryck, Decru, and Vercauteren in 2020 [17]. The authors showed that using radical isogeny formulas in CSIDH-512 leads to more efficient implementation and a speed-up of 19%, see [17, Section 6.]. Furthermore, in 2022, the same group of authors, along with Houben [16], proposed new and optimized radical isogeny formulas and achieved a speed-up of 35% of CSIDH-512 compared to the implementation without radical isogenies. In [17], formulas were given for $N \leq 13$, and in [16] authors provided formulas for $N \leq 37$.

The concept of radical isogeny formulas was initially introduced for elliptic curves in Tate normal form. Generally, an elliptic curve over a field k and a point on that curve of an order N, where $N \geq 4$, are isomorphic to a unique pair of an elliptic curve of the form

$$E\colon y^2 + (1-c)xy - by = x^3 - bx^2 \quad \text{with} \quad b, c \in k,$$

and a point $P = (0,0)$ of the same order N. This form is known as the Tate normal form and it provides two unique coefficients, denoted b and c. Given a cyclic isogeny $\varphi\colon E \to E' = E/\langle P \rangle$, radical isogeny formulas compute points P' of order N on E' such that the composition

$$E \xrightarrow{\varphi} E' \to E'/\langle P' \rangle$$

is cyclic of degree N^2. The coordinates of P' are elements of the smallest field that contains the coefficients b and c, along with a radicand ρ that is a N-th root of a rational expression in the coefficients b and c. The elliptic curve E' and point P' are also isomor-

phic to an elliptic curve in Tate normal form (for example, defined with coefficients b' and c') and a point $(0,0)$ of order N. This allows us to use radical isogeny formulas again, making the process iterative. The coefficients b' and c' can be expressed as elements of the same field as P'.

As a first contribution of this thesis, in Chapter 4, we will extend the notion of radical isogeny formulas to the language of modular curves. To achieve this, we will utilize enhanced elliptic curves, which are curves paired with additional torsion data and affiliated with some congruence subgroup. The aforementioned parameters from Tate normal form and the radicand ρ can all be regarded as functions on the set of equivalence classes of enhanced elliptic curves. This generalization of radical isogenies for degree N is directly related to the modular curve $X_1(N)$, congruence subgroup $\Gamma_1(N)$, and pairs of enhanced elliptic curves consisting of an elliptic curve and a point of order N.

In [54], Onuki and Moriya introduced radical isogeny formulas of degrees 3 and 4 on Montgomery curves. A Montgomery curve over a field k is an elliptic curve of the form

$$E\colon By^2 = x^3 + Ax^2 + x,$$

where $A, B \in k$ and $B(A^2-4) \neq 0$. Typically, the value of B is set to 1. The coefficient A is called the Montgomery coefficient of E. For degree 4 (degree 3 is similar), there exists a bijection between the set of equivalence classes of enhanced elliptic curves for $\Gamma_0(4)$, denoted by $S_0(4)$, and the set of equivalence classes of enhanced elliptic curves for $\Gamma_1(4)$. This bijective relation implies the existence of radical isogeny formulas for the modular curve $X_0(4)$. The Montgomery coefficient A represents a class in the set $S_0(4)$, see [54, Section 2.3.]. In other words, the coefficient A describes an enhanced elliptic curve where the additional torsion data is a cyclic subgroup of order 4. The Montgomery coefficient for the curve E' can be calculated by a rational expression depending on the fourth root from $4(A+2)$, see [54, Theorem 8.].

The authors of [54] explored the possibility of extending radical isogeny formulas to the modular curve $X_0(N)$ when $N \geq 5$. We can summarize the idea behind this in a few informal stages: take a modular curve of genus zero, such as $X_0(5)$. Find a parameter that specifies its set of equivalence classes of enhanced elliptic curves, then a model of a elliptic curve over $X_0(5)$ defined by that parameter (Tate, Montgomery, or something

else), and then find a radical isogeny formula on such a curve. This approach is presented as an example in [54, Section 4.] and in Example 3.2.5, which argues against the existence of radical isogeny formulas for that curve. While it suggests that finding radical isogenies for degrees greater than 4 may not be possible, a general answer was left as an open problem. We provide a solution to this open problem in Chapter 5 by proving radical isogeny formulas cannot exist on the set of equivalence classes of enhanced elliptic curves for $\Gamma_0(N)$ when $N \geq 5$. To achieve this, we extend the setting of the previous chapter to include modular curve $X_0(N)$.

Chapter 1 presents the necessary background, notation, definitions, and results that will be used throughout this thesis. This includes a brief overview of some parts of group theory, fields, fields extension, and Galois theory. We will also cover elliptic curves, isogenies of elliptic curves, divisors, pairings, congruence subgroups, and modular curves.

Chapter 2 reviews post-quantum cryptography, mainly focusing on isogeny-based post-quantum cryptography. We will provide a brief introduction to cryptography and elliptic-curve cryptography. We will also introduce basic primitives used in isogeny-based post-quantum cryptography, and give an overview of CSIDH key exchange.

Chapter 3 provides the background on radical isogenies, including the definition, some of their properties, and examples. This chapter will also introduce radical isogenies on Montgomery curves and delve deeper into the previously mentioned open problem.

1. PRELIMINARIES

This chapter provides the necessary background. For the most part, the section regarding some basic properties of groups, fields, and Galois theory follows [46]. The elliptic curves section mostly follows [64] and [69], and the modular curves section follows [31].

1.1. GROUPS, GROUPS ACTION AND OVERVIEW OF GALOIS THEORY

This section provides some basic definitions regarding the theory of groups, groups action, and a short overview of basic notions from Galois theory. We will begin by defining a group.

Definition 1.1.1. A group G is a set, together with a binary operation $\cdot\colon G\times G\to G$, that maps a pair $(x,y)\in G\times G$ to an element of G denoted $x\cdot y$, and has the following properties:

(i) It is associative:

$$(x\cdot y)\cdot z = x\cdot(y\cdot z),\quad \forall x,y,z\in G.$$

(ii) It has a unique unit element (identity):

$$\exists!\ e\in G \quad \text{such that} \quad e\cdot x = x = x\cdot e,\quad \forall x\in G.$$

(iii) Every element has a unique inverse:

$$\forall x\in G,\ \exists!\ y\in G \quad \text{such that} \quad x\cdot y = y\cdot x = e.$$

Notice that the existence of the unit element also implies that G is a nonempty set. The usual notation for the operation $x \cdot y$ is just xy, and for the inverse, it is x^{-1}. When the underlying operation is addition, we use the $+$ notation and $-x$ for the inverse. If the binary operation is also commutative, G is called commutative or abelian group.

Example 1.1.2. The set of rational numbers forms a group under addition, denoted $(\mathbb{Q}, +)$. The set of nonzero rational numbers forms a group under multiplication, denoted $(\mathbb{Q}^{\times}, \cdot)$.

Definition 1.1.3. A group $(G, \cdot)$ is a cyclic group if there exists an element $g \in G$ such that every element $x \in G$ can be written in a form $x = g^n$ for some integer n. The element g is called a generator of the group G.

Let G and G' be groups. A homomorphism between G and G' is a function $f\colon G \to G'$ such that $f(xy) = f(x)f(y)$, for all $x, y \in G$. Additionally, f maps the unit element e of G to the unit element e' of G'. A homomorphism between G and G' is an isomorphism if there exists a homomorphism $g\colon G' \to G$ such that $f \circ g$ and $g \circ f$ are the identity mappings on G' and G respectively. A homomorphism $f\colon G \to G$ is called an endomorphism. An isomorphism $f\colon G \to G$ is called an automorphism.

Example 1.1.4. The group $(\mathbb{Z}, +)$ is a cyclic group with a generator 1 (and also -1). Any infinite cyclic group is isomorphic to this group. The group $(\mathbb{Z}/N\mathbb{Z}, +)$ is a cyclic group with a generator 1. Any cyclic group with N elements is isomorphic to this group.

Let $(G, \cdot)$ be a group. A subgroup H of G is a subset of G containing the unit element of G, closed under the inverse, and closed under the binary operation of G. In other words, H is a subgroup of G if the restriction of the operation $\cdot$ to $H \times H$ is a group operation on H.

Example 1.1.5. Let G, G' be groups and let f be a homomorphism from G to G'. The kernel of the map f is a subset of G containing all elements x of G, such that $f(x) = e'$, where e' is a unit element for the group G'. The kernel of the map f is a subgroup of G.

Let $(G, \cdot)$ be a group, and let H be a subgroup of G. A subgroup H is a normal subgroup if

$$\forall x \in G \quad \text{we have} \quad xHx^{-1} = H.$$

The kernel of the homomorphism f from Example 1.1.5 is a normal subgroup.

Definition 1.1.6. Let $(G,\cdot)$ be a group and let S be a set. The (left) group action of G on S is a mapping $G\times S\to S$ satisfying:

(i) If e is the unit element of G, then $es=s$, for all $s\in S$.

(ii) For all $x,y\in G$ and $s\in S$, we have $x(ys)=(xy)s$.

A group action $G\times S\to S$ is free if for all $s\in S,\ gs=s$ implies $g=e$. In other words, only the identity element fixes elements from the set S. The subset of S consisting of all elements xs for $x\in G$, is called the orbit of s under G and it is denoted by Gs.

Galois theory

A number of results of this thesis are closely related to the notion of Galois extension. Despite Galois theory being too broad of a subject to be covered in such a short section, for completeness we will introduce the basic definitions and results, starting with the definition of a field and a field extension.

Definition 1.1.7. Let k be a set together with two binary operations, multiplication and addition. Then $(k,+,\cdot)$ is a field if:

(i) With respect to addition, k is a commutative group, where the unit element is denoted by 0.

(ii) Multiplication is associative and commutative, and the unit element is denoted by 1.

(iii) $0\neq 1$, and every nonzero element has a multiplicative inverse.

(iv) (distributivity) For all $x,y,z\in k$,

$$(x+y)z=xz+yz \quad \text{and} \quad z(x+y)=zx+zy.$$

Note that every field has at least two elements. The field k is finite if it contains a finite number of elements. The number of elements in a finite field is also the order of that field.

Example 1.1.8. The real numbers $\mathbb{R}$, with the usual operations of addition and multiplication form a field. The group of integers modulo prime number p, denoted by $\mathbb{Z}/p\mathbb{Z}$ or $\mathbb{F}_p$, is a finite field.

The characteristic of a field k, denoted by $\text{char}(k)$, is the smallest positive integer n such that

$$\underbrace{1+\cdots+1}_{n \text{ times}} = 0,$$

i.e. the smallest positive integer needed to add multiplicative identity to get the additive identity. If such an n does not exist, then the characteristic of the field is equal to zero.

Definition 1.1.9. A field k is a subfield of a field L if $k \subseteq L$ and field operations in k are inherited from L.

The characteristic of a subfield is the same as the characteristics of the larger field. If k is a subfield of a field L, then we also say that L is an extension field of k. This is usually denoted L/k.

Example 1.1.10. The field of complex numbers $\mathbb{C}$ is a field extension of the field of real numbers $\mathbb{R}$.

There are several different types of field extensions, but we will consider only those required to define a Galois extension.

Definition 1.1.11. Let k be a subfield of a field L. An element α of L is algebraic over k if there exist elements $a_0, \ldots, a_n$, $n \geq 1$ in k, not all equal to zero, such that

$$a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0.$$

To put it differently, an element of L is algebraic if it is a root of a nonzero polynomial with coefficients in k. The set of algebraic elements of L over k is called the algebraic closure of k, usually denoted by $\overline{k}$. A field extension L of k is an algebraic extension if every element of L is algebraic over k.

Definition 1.1.12. An algebraic extension L/k is a normal field extension if every irreducible polynomial of $k[x]$ which has a root in L splits into linear factors in L.

Definition 1.1.13. An algebraic extension L/k is a separable extension if the minimal polynomial of every element of L is separable (does not have multiple roots in an algebraic closure over k).

Now we can define a Galois extension.

Definition 1.1.14. An algebraic extension L of a field k is a Galois extension if it is normal and separable.

Let L/k be a field extension. Automorphism σ of L such that $\sigma(x) = x$, for all $x \in k$, is called a k-automorphism of L. A set of all k-automorphism of L is denoted by $\mathrm{Aut}(L/k)$. This, with respect to composition, is the automorphism group of L/k.

Definition 1.1.15. Let k be a subfield of a field L, and H a subgroup of automorphisms of $\mathrm{Aut}(L/k)$. The fixed field of H is defined by

$$L^H = \{l \in L \colon \sigma(l) = l,\ \forall \sigma \in H\}.$$

Remark. If L/k is an algebraic extension, then L/k is Galois extension if $L^{\mathrm{Aut}(L/k)} = k$.

The group of automorphisms of a Galois extension L over k is called the Galois group of L over k and is usually denoted by $\mathrm{Gal}(L/k)$. The identity element of this group is the identity function on L. A cyclic extension is a Galois extension whose Galois group is cyclic.

The main result of the Galois theory is that the Galois extension and Galois group can be used to describe intermediate fields. This is given in the fundamental theorem of Galois theory.

Theorem 1.1.16. Let L be a finite Galois extension of k, with Galois group G, and let $\sigma \in G$. There is a bijection between the set of subfields E of L containing k (E is an intermediate field), and the set of subgroups H of G, given by $E = L^H$. The field E is Galois over k if and only if H is normal in G, and if that is the case, then the map $\sigma \mapsto \sigma|_E$ induces an isomorphism of G/H onto the Galois group of E over k.

Proof. See [46, Chapter VI, Theorem 1.1.] ■

We will finish this section with a definition of radical and simple radical extension, the importance of which will become obvious in Chapter 3.

Definition 1.1.17. A radical field extension L of k is a field extension obtained by adjoining a sequence of n-th roots of elements.

Definition 1.1.18. Let k be a field. Field extension L/k is a simple radical extension of degree $N \geq 2$ if there exists $\alpha \in L$ such that:

(i) $L = k(\alpha)$,

(ii) $\alpha^N \in k$,

(iii) $x^N - \alpha^N \in k[x]$ is irreducible.

Example 1.1.19. Examples of radical extensions of $\mathbb{Q}$ are $\mathbb{Q}(\sqrt{2})$, $\mathbb{Q}(\sqrt[3]{2})$ or more generally $\mathbb{Q}(\sqrt[n]{2})$.

1.2. ELLIPTIC CURVES

This section provides some basic definitions and results regarding the theory of elliptic curves, isogenies of elliptic curves, divisors, and pairings. We start with the definition of an elliptic curve.

Definition 1.2.1. Let k be a field. An elliptic curve E over k is a smooth (nonsingular) projective curve of genus one with a specified base point $\mathcal{O}_E$. The point $\mathcal{O}_E$ is called the point at infinity.

Every such curve has an equation of the form

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3,$$

where $\mathcal{O}_E = [0,1,0]$ is the point at infinity and coefficients $a_1, a_2, a_3, a_4, a_6 \in k$. Using the transformations $x = \frac{X}{Z}$ and $y = \frac{Y}{Z}$ we get the usual form of elliptic curves called the long Weierstrass equation

$$E\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{1.1}$$

always keeping in mind that there is an extra point at infinity.

In special cases, when the field characteristic is not equal to 2 or 3, the long Weierestrass equation can be reduced to the short Weierstrass equation. Let $\mathrm{char}(k) \neq 2$. The substitution

$$y \mapsto \frac{1}{2}(y - a_1x - a_3)$$

transforms the equation (1.1) into

$$E\colon y^2 = 4x^3 + b_2x^2 + 2b^4x + b_6,$$

where

$$b_2 = a_1^2 + 4a_4, \quad b_4 = 2a_4 + a_1a_3, \quad b_6 = a_3^2 + 4a_6.$$

If additionally, $\mathrm{char}(k) \neq 3$, using the substitution

$$(x,y) \mapsto \left(\frac{x - 3b_2}{36}, \frac{y}{108}\right)$$

we get the short Weierstrass equation

$$y^2 = x^3 + 27c_4x - 54c_6,$$

where $c_4 = b_2^2 - 24b_4$ and $c_6 = -b_2^3 + 36b_2b_4 - 216b_6$. Let b_8 denote the coefficient

$$b_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2.$$

With coefficients a_i, b_i, c_i we can define the quantity Δ, which is the discriminant of the elliptic curve E equal to

$$\Delta(E) = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6,$$

and j-invariant equal to

$$j(E) = \frac{c_4^3}{\Delta(E)}.$$

When E is given as $y^2 = x^3 + ax + b$, discriminant and j-invariant are equal to

$$\Delta(E) = -16(4a^3 + 27b^2), \quad j(E) = -1728\frac{(4a)^3}{\Delta(E)}.$$

The smoothness (nonsingularity) of the elliptic curve is generally characterized by the values of partial derivations of the function at the points on the curve, i.e. at least one partial derivation in a point should be different from zero. When a curve is given as $E\colon y^2 = x^3 + ax + b$, nonsigularity is characterized by $\Delta(E) \neq 0$.

One of the most important properties of an elliptic curve is the ability to build a group law on the points of the curve.

Definition 1.2.2 (Composition law, see [64, Chapter III, Section 2.])**.** Let E be an elliptic curve over a field k. Let $P, Q \in E$, let l be the line through P and Q (if $P = Q$, let l be the tangent line to E at P), and let R be the third point of intersection of l with E. Let l' be the line through R and $\mathcal{O}_E$. Then l' intersects E at $R, \mathcal{O}_E$, and a third point. We denote that third point by $P + Q$.

The composition law has all the expected "addition" properties: it is commutative and associative, it has a neutral element and an inverse; and, if we add three points on the same line we get a neutral element.

Proposition 1.2.3. Together with the Composition law from Definition 1.2.2, E is an abelian group with identity element $\mathcal{O}_E$. Additionally, suppose that E is defined over k. Then

$$E(k) = \{(x,y) \in k \times k \colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6\} \cup \{\mathcal{O}_E\},$$

a set of all the points on E defined over k, is a subgroup of $E(\overline{k})$.

Proof. See [64, Chapter III, Proposition 2.2.]. ■

Explicit formulas for the addition of points on a curve (1.1) are given in [64, Chapter III, Section 2.2.]. The following theorem gives the formula for an elliptic curve in the short Weierstrass form.

Theorem 1.2.4. Let E be an elliptic curve over a field k of the form

$$E \colon y^2 = x^3 + ax + b,$$

and let P and Q be points on E.

(i) If $P = \mathcal{O}_E$, then $P + Q = Q$.

(ii) If $Q = \mathcal{O}_E$, then $P + Q = P$.

(iii) If $P,\ Q \neq \mathcal{O}_E$, write $P = (x_1, y_1)$ and $Q = (x_2, y_2)$.

(iv) If $x_1 = x_2$ and $y_1 = -y_2$, then $P + Q = \mathcal{O}_E$.

(v) Otherwise, let λ be defined as

$$\lambda = \begin{cases} \frac{y_2 - y_1}{x_2 - x_1} & \text{if } P \neq Q, \\ \frac{3x_1^2 + a}{2y_1} & \text{if } P = Q, \end{cases}$$

and let

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

Then $P + Q = (x_3, y_3)$.

Proof. See [65, Chapter 6, Theorem 6.6.] ■

Definition 1.2.5. Let E be an elliptic curve, $P \in E$ a point and $N \in \mathbb{Z}$. A multiplication by N map is a map $[N] \colon E \to E$ defined with:

- $[N]P = \overbrace{P + \cdots + P}^{N \text{ terms if } N>0}$,
- $[N]P = \overbrace{-P - \cdots - P}^{|N| \text{ terms if } N<0}$,
- $[0]P = \mathcal{O}_E$.

A point P on the elliptic curve E is of order N if

$$[N]P = \mathcal{O}_E \quad \text{and} \quad [m]P \neq \mathcal{O}_E \quad \text{for} \quad 0 < m < N.$$

Closely related to the map $[N]$ is the torsion subgroup.

Definition 1.2.6. Let E be an elliptic curve and let $N \in \mathbb{Z}$ with $N \geq 1$. The N-torsion subgroup of E, denoted by $E[N]$ is the set

$$E[N] = \{P \in E(\overline{k}) \colon [N]P = \mathcal{O}_E\}.$$

The torsion subgroup of E, denoted by E_{tors}, is a set of points of finite order,

$$E_{tors} = \bigcup_{N=1}^{\infty} E[N].$$

If E is defined over k, then $E_{tors}(k)$ denotes the points of finite order in $E(k)$.

The following lemma gives the structure of the torsion subgroup.

Lemma 1.2.7. Let E be an elliptic curve over a field k and let $N \in \mathbb{Z}$ with $N \geq 1$. The N-torsion group of E has the following structure:

- If $\text{char}(k) \nmid N$:

$$E[N] \simeq (\mathbb{Z}/N\mathbb{Z})^2.$$

- If $\text{char}(k) = p > 0$, then one of the following is true:

  (i) $E[p^e] = \frac{\mathbb{Z}}{p^e\mathbb{Z}}$ for all $e \geq 1$,

  (ii) $E[p^e] = \{\mathcal{O}_E\}$ for all $e \geq 1$.

Proof. See [64, Chapter III, Corollary 6.4.]. ■

In the second case in Lemma 1.2.7, for curves defined over a field of positive characteristic p, the elliptic curve is ordinary when the torsion group is equal to $E[p] \simeq \mathbb{Z}/p\mathbb{Z}$, and it is supersingular when the torsion group is equal to $E[p] \simeq \{\mathcal{O}_E\}$.

The following lemma holds for a curve E as in (1.1), and a point P of order $N \geq 4$.

Lemma 1.2.8. Let E be an elliptic curve over k and let $P \in E(k)$ be a point of order $N \geq 4$. The pair (E,P) is isomorphic to a unique pair $(\overline{E},\overline{P})$ of the form

$$\overline{E}\colon y^2 + (1-c)xy - by = x^3 - bx^2,\ \overline{P} = (0,0) \tag{1.2}$$

with $b,c \in k$ and

$$\Delta(b,c) = b^3(c^4 - 8bc^2 - 3c^3 + 16b^2 - 20bc + 3c^2 + b - c) \neq 0.$$

Proof. This lemma will be used several times throughout this thesis. We give a complete proof, and for additional details we refer to [66, Lemma 2.1.] and [35, Section 15.3.].

Let $P = (x_P, y_P)$ be a point of order $N \geq 4$ on an elliptic curve

$$E\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

Assume that P is equal to $(0,0)$, if not, use the translation $(x,y) \mapsto (x - x_P, y - y_P)$. This implies $a_6 = 0$, and, as the curve is nonsingular, one of the coefficients a_3 and a_4 should be nonzero. The point P is not a point of order 2, so we can assume that $a_4 = 0$, meaning $a_3 \neq 0$ and the substitution $(x,y) \mapsto (x, y + \frac{a_4}{a_3}x)$ does not affect $(0,0)$. The point P is also not a point of order 3, so $a_2 \neq 0$. Set $u = \frac{a_2}{a_3}$. Substitution $(x,y) \mapsto \left(\frac{x}{u^2}, \frac{y}{u^3}\right)$ does not affect the point $P = (0,0)$, and the curve equation is equal to

$$y^2 + a_3^{-1}a_1a_2xy + a_3^{-2}a_2^3y = x^3 + a_3^{-2}a_2^3x^2.$$

To get the isomorphic curve $\overline{E}$, set $b = -a_3^{-2}a_2^3$ and $c = 1 - a_3^{-1}a_1a_2$. The equation for curve $\overline{E}$ is equal to

$$\overline{E}\colon y^2 + (1-c)xy - by = x^3 - bx^2.$$

The uniqueness of the parameters b and c, i.e. of the curve $\overline{E}$, follows from the standard change of coordinates in [64, Chapter III, Table 3.1.] and [64, Chapter III, Proposition 3.1.(b).].

The quantity $\Delta(b,c)$ is the discriminant of $\overline{E}$ so it is not zero. Conversely, if $\Delta(b,c)$ is nonzero, then $(\overline{E},\overline{P})$ defines an elliptic curve and a point that does not have order $1,2$ or 3. ■

The curve $\overline{E}$ in (1.2) is said to be in Tate normal form.

Division polynomials are a useful tool for calculating multiples of points on an elliptic curve.

Definition 1.2.9. Let E be an elliptic curve given by a Weierstrass equation

$$E\colon y^2+a_1xy+a_3y=x^3+a_2x^2+a_4x+a_6,$$

and let $m\in\mathbb{Z}$. The m-division polynomial $\psi_m\in\mathbb{Z}[a_1,\ldots,a_6,x,y]$ is given by

$$\begin{aligned}
\psi_1&=1,\\
\psi_2&=2y+a_1x+a_3,\\
\psi_3&=3x^4+b_2x^3+3b_4x^2+3b_6x+b_8,\\
\psi_4&=\psi_2\cdot\left(2x^6+b_2x^5+5b_4x^4+10b_6x^3+10b_8x^2+(b_2b_8-b_4b_6)x+(b_4b_8-b_6^2)\right),
\end{aligned}$$

and then inductively by

$$\begin{aligned}
\psi_{2m+1}&=\psi_{m+2}\psi_m^3-\psi_{m-1}\psi_{m+1}^3 && \text{for } m\geq 2,\\
\psi_2\psi_{2m}&=\psi_{m-1}^2\psi_m\psi_{m+2}-\psi_{m-2}\psi_m\psi_{m+1}^2 && \text{for } m\geq 3.
\end{aligned}$$

Additionally, we define two more polynomials,

$$\begin{aligned}
\phi_m&=x\psi_m^2-\psi_{m+1}\psi_{m-1},\\
4y\omega_m&=\psi_{m-1}^2\psi_{m+2}+\psi_{m-2}\psi_{m+1}^2.
\end{aligned}$$

Let $P=(x_1,y_1)$ be a point on an elliptic curve E. Then, $[N]P$ can be calculated as

$$[N]P=\left(\frac{\phi_N(P)}{\psi_N(P)^2},\frac{\omega_N(P)}{\psi_N(P)^3}\right). \tag{1.3}$$

Thus, P is a point of order dividing N on the curve if and only if $\psi_N(P)=0$. The same definition of division polynomials can be used for the form (1.2). Let $P_m\in\mathbb{Z}[b,c]$ be the m-th division polynomial of the curve (1.2) evaluated at $(0,0)$. If $N\geq 4$ and (E,P) is as

in Lemma 1.2.8, i.e. we have unique $b,c \in k$ with $\Delta(b,c) \neq 0$, then the point $P \in E(K)$ has order dividing N if and only if $P_N(b,c) = 0$. The first couple of polynomials P_m are

$$\begin{aligned} P_1(b,c) &= 1, \\ P_2(b,c) &= -b, \\ P_3(b,c) &= -b^3, \\ P_4(b,c) &= c \cdot b^5, \\ P_5(b,c) &= -(c-b) \cdot b^8. \end{aligned}$$

For $m \geq 4$, let $F_m \in \mathbb{Z}[B,C]$ be P_m, with all the factors in common with $\Delta(b,c)$ and P_d, for $d < m$, removed. The first couple of polynomials F_m are

$$\begin{aligned} F_2(b,c) &= \frac{b^4}{\Delta(b,c)} \in \mathbb{Q}(b,c), \\ F_3(b,c) &= b, \\ F_4(b,c) &= c, \\ F_5(b,c) &= c-b. \end{aligned} \tag{1.4}$$

For $N \geq 4$, the point $P = (0,0)$ on E as in (1.2) is of order N if and only if $F_N = 0$. For more details, see [66]. We will use polynomials F_N in Chapter 3.

1.2.1. Isogenies of elliptic curves

This section covers the basics of isogenies of elliptic curves. For a general introduction to the algebraic maps between projective varieties, we refer to [64, Chapter I, Section 3]. We give a short excerpt.

Definition 1.2.10. Let E and E' be elliptic curves over a field k. The map $\varphi\colon E \to E'$ is a rational map if it is defined as $\varphi = (u,v)$, where u and v are rational functions in $k(E)$, and not both equal to zero. For a point $P \in E(\bar{k})$, we have $\varphi(P) = (u(P), v(P))$. The map φ is defined at a point P if there exists a function $g \in k(C)^*$ such that $u \circ g$ and $v \circ g$ are defined at P.[1] The map φ is a morphism if it is defined for every point $P \in E(\bar{k})$.

Two elliptic curves E/k and E'/k are isomorphic over $\bar{k}$ if there exists morphisms $\varphi\colon E \to E'$ and $\phi\colon E' \to E$ such that $\varphi \circ \phi$ and $\phi \circ \varphi$ are identity maps on E' and E

[1] Definition 1.2.24 outlines the conditions for a function to be considered defined at a point.

respectively. The previously defined j-invariant provides a way to check if two elliptic curves are isomorphic.

Proposition 1.2.11. Two elliptic curves are isomorphic over $\overline{k}$ if and only if they have the same j-invariant.

Proof. See [64, Chapter III, Proposition 1.4.(b).]. ■

Theorem 1.2.12. Let φ be a morphism of curves. Then φ is either constant or surjective.

Proof. See [64, Chapter II, Theorem 2.3.]. ■

Let E and E' be curves over a field k and let φ be a nonconstant rational map defined over k. Composition with φ induces an injection of function fields (also called pullback) that fixes the field k,

$$\varphi^*\colon k(E') \to k(E), \quad \varphi^* f \mapsto f \circ \varphi.$$

The function field $k(E)$ is a finite extension of $\varphi^*(k(E'))$.

Definition 1.2.13. Let E and E' be elliptic curves. An isogeny $\varphi\colon E \to E'$ is a nonconstant morphism satisfying $\varphi(\mathcal{O}_E) = \mathcal{O}_{E'}$.

Two elliptic curves E and E' are isogenous if there is an isogeny from E to E'. Being isogenous is an equivalence relation.

Example 1.2.14. The previously defined multiplication by N map is an isogeny.

Definition 1.2.15. Let $\varphi\colon E \to E'$ be a map of curves defined over k. If φ is constant, we define the degree of φ to be 0. Otherwise, we say that φ is a finite map and we define its degree to be

$$\deg(\varphi) = [k(E)\colon \varphi^*(k(E'))].$$

The degree of the multiplication by N map is N^2. The degree of the zero isogeny is zero. An isogeny is separable (inseparable, purely inseparable) if the finite extension $k(E)/\varphi^*(k(E'))$ is separable (inseparable, purely inseparable).

Theorem 1.2.16. Let $\varphi\colon E \to E'$ be an isogeny. Then φ is a homomorphism, i.e.

$$\varphi(P+Q) = \varphi(P) + \varphi(Q), \quad \forall P,\, Q \in E.$$

Proof. See [64, Chapter III, Theorem 4.8.]. ■

As a direct consequence of Theorem 1.2.16, we have that the kernel of an isogeny φ is a finite subgroup $\ker(\varphi) = \varphi^{-1}(\mathcal{O}_{E'})$ of $E(\overline{k})$. The size of the kernel divides the degree of the isogeny, and they are equal when the isogeny is separable. An isogeny is cyclic if its kernel is a cyclic group.

Proposition 1.2.17. Let E be an elliptic curve and let C be a finite subgroup of E. There exists a unique elliptic curve E' and a separable isogeny $\varphi\colon E \to E'$ satisfying $\ker(\varphi) = C$.

Proof. See [64, Chapter III, Theorem 4.12.]. ■

Proposition 1.2.17 implies the existence of an isogeny from a finite subgroup of a curve. The proof of the Proposition 1.2.17 given in [64] is not constructive but there is also a way to give a more algebraic proof with explicit formulas to build such an isogeny.

Theorem 1.2.18. Let E be an elliptic curve over a field k, given by the Weierstrass equation

$$E\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \quad a_i \in k, \forall i.$$

Let C be a finite subgroup of $E(\overline{k})$. Then there exists an elliptic curve E' and a separable isogeny φ from E to E' such that $C = \ker(\varphi)$. For a point $Q = (x_Q, y_Q) \in C$ with $Q \neq \mathcal{O}_E$, define

$$\begin{aligned}
g_Q^x &= 3x_Q^2 + 2a_2x_Q + a_4 - a_1y_Q,\\
g_Q^y &= -2y_Q - a_1x_Q - a_3,\\
v_Q &= \begin{cases} g_Q^x, & \text{if } 2Q = \mathcal{O}_E,\\ 2g_Q^x - a_1g_Q^y, & \text{else} \end{cases}\\
u_Q &= (g_Q^y)^2.
\end{aligned}$$

Let C_2 be the points of order 2 in C. Choose $R \subset C$ such that we have a disjoint union

$$C = \{\mathcal{O}_E\} \cup C_2 \cup R \cup (-R)$$

(in other words, for each pair of non-2-torsion points P, $-P \in C$, put exactly one of them in R). Let $S = R \cup C_2$. Set

$$v = \sum_{Q \in S} v_Q, \quad w = \sum_{Q \in S} (u_Q + x_Q v_Q).$$

Then E' has the equation

$$Y^2 + A_1XY + A_3Y = X^3 + A_2X^2 + A_4X + A_6,$$

where

$$A_1 = a_1,\ A_2 = a_2,\ A_3 = a_3,\ A_4 = a_4 - 5v,\ A_6 = a_6 - (a_1^2 + 4a_2)v - 7w.$$

The isogeny is given by

$$X = x + \sum_{Q \in S} \left(\frac{v_Q}{x - x_Q} + \frac{u_Q}{(x - x_Q)^2} \right),$$

$$Y = y - \sum_{Q \in S} \left(u_Q \frac{2y + a_1x + a_3}{(x - x_Q)^3} + v_Q \frac{a_1(x - x_Q) + y - y_Q}{(x - x_Q)^2} + \frac{a_1u_Q - g_Q^x g_Q^y}{(x - x_Q)^2} \right).$$

Proof. See [69, Chapter 12, Theorem 12.16.]. ■

The formulas from Theorem 1.2.18 were originally given in [67] by Vélu, hence they are usually called Vélu's formulas. These formulas are the main tool in calculating isogenies between elliptic curves.

Theorem 1.2.19. Let $\varphi\colon E \to E'$ be an isogeny of degree N. There exists a unique isogeny

$$\widehat{\varphi}\colon E' \to E \quad \text{such that} \quad \widehat{\varphi} \circ \varphi = [\deg(\varphi)].$$

Proof. See [64, Chapter III, Theorem 6.1.]. ■

The isogeny from Theorem 1.2.19 is called the dual isogeny. The following theorem addresses some properties of the dual isogeny.

Theorem 1.2.20. Let $\varphi\colon E \to E'$ be an isogeny of degree N. Then:

(i) $\widehat{\varphi} \circ \varphi = [N]$ on E and $\varphi \circ \widehat{\varphi} = [N]$ on E'.

(ii) Let $\lambda : E' \to E''$ be an isogeny. Then

$$\widehat{\lambda \circ \varphi} = \widehat{\varphi} \circ \widehat{\lambda}.$$

(iii) Let $\psi\colon E \to E'$ be an isogeny. Then

$$\widehat{\varphi + \psi} = \widehat{\varphi} + \widehat{\psi}.$$

(iv) For all $m \in \mathbb{Z}$,

$$\widehat{[m]} = [m] \quad \text{and} \quad \deg([m]) = m^2.$$

(v) $\deg(\widehat{\varphi}) = \deg(\varphi)$.

(vi) $\widehat{\widehat{\varphi}} = \varphi$.

Proof. See [64, Chapter III, Theorem 6.2.]. ■

1.2.2. Divisors, Weil, and Tate pairing

We will start by defining a divisor for a curve and several other related concepts.

Definition 1.2.21. Let E be a curve. A divisor group for E, denoted by $\mathrm{Div}(E)$, is a free abelian group generated by the points of E. A divisor $D \in \mathrm{Div}(E)$ is defined as a formal sum

$$D = \left\{ \sum_{P \in E} n_P(P) \colon n_P \in \mathbb{Z},\ n_P = 0, \text{ for all but finitely many } P \in E \right\}.$$

The degree of D is defined by

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

The support of a divisor is the set of points $P \in E$ for which $n_P \neq 0$. The divisors of degree equal to zero form a subgroup of $\mathrm{Div}(E)$:

$$\mathrm{Div}^0(E) = \{D \in \mathrm{Div}(E) \colon \deg(D) = 0\}.$$

If a curve E is defined over a field k, the Galois group $\mathrm{Gal}(\bar{k}/k)$ acts on $\mathrm{Div}(E)$ and $\mathrm{Div}^0(E)$ as

$$D^\sigma = \sum_{P \in E} n_P(P^\sigma), \quad \sigma \in \mathrm{Gal}(\bar{k}/k).$$

A divisor D is defined over k if $D^\sigma = D$, for all $\sigma \in \mathrm{Gal}(\bar{k}/k)$. Divisors defined over k are called k-rational divisors and they form a group denoted by $\mathrm{Div}_k(E)$.

Building on the definition of the divisor of a curve, we will now define the divisor for a function $f \in \bar{k}(E)^*$. This requires additional theoretical background and definitions that we will provide here, but not in full detail. For more, refer to [64, Chapter II, Section 1.]. We will assume that the reader is familiar with the definitions of a ring, an ideal, and a maximal ideal. Otherwise, they are available in [46].

Proposition 1.2.22. Let E be a curve and $P \in E$ a smooth point. Then $\overline{k}[E]_P$ (the local ring of E in P) is a discrete valuation ring (a principal ideal domain with unique maximal ideal, but not a field).

Proof. See [64, Chapter II, Proposition 1.1.]. ■

Definition 1.2.23. Let E be a curve and $P \in E$ a smooth point. The (normalized) valuation on $\overline{k}[E]_P$ is given by

$$\mathrm{ord}_P \colon \overline{k}[E]_P \to \{0,1,2,\ldots\} \cup \{\infty\}, \quad \mathrm{ord}_P(f) = \sup\{d \in \mathbb{Z} \colon f \in M_p^d\},$$

where M_p is the maximal ideal of $\overline{k}[E]_P$.

Using $\mathrm{ord}_P(f/g) = \mathrm{ord}_p(f) - \mathrm{ord}_p(g)$, it is possible to extend ord_P to $\overline{k}(E)$,

$$\mathrm{ord}_P \colon \overline{k}(E) \to \mathbb{Z} \cup \{\infty\}.$$

A uniformizer for E at a point P is any function $t \in \overline{k}(E)$ with $\mathrm{ord}_P(t) = 1$, i.e. a generator for the ideal M_P.

Definition 1.2.24. Let E be a curve, $P \in E$ a smooth point and $f \in \overline{k}(E)^*$. The order of f at P is $\mathrm{ord}_P(f)$. If $\mathrm{ord}_P(f) > 0$, then f has a zero at P, and if $\mathrm{ord}_P(f) < 0$, then f has a pole at P. If $\mathrm{ord}_P(f) \geq 0$, then f is regular (or defined) at P and we can evaluate $f(P)$. Otherwise, f has a pole at P and we write $f(P) = \infty$.

Proposition 1.2.25. Let E be a smooth curve and $f \in \overline{k}(E)$ with $f \neq 0$. There are only finitely many points of the curve E at which f has a pole or zero. Furthermore, if f has no poles, then $f \in \overline{k}$, i.e. it is a constant.

Proof. See [64, Chapter II, Proposition 1.2.]. ■

Now we are ready to define a divisor of a function. Let E be a smooth curve and $f \in \overline{k}(E)^*$. The divisor of the function f is given by

$$\mathrm{div}(f) = \sum_{P \in E} \mathrm{ord}_P(f)(P).$$

This is well-defined because of Proposition 1.2.25. If $\sigma \in \mathrm{Gal}(\overline{k}/k)$, then

$$\mathrm{div}(f^\sigma) = (\mathrm{div}(f))^\sigma.$$

Furthermore, if $f \in k(E)$, then $\mathrm{div}(f) \in \mathrm{Div}_k(E)$. The divisor of a function is used to define a principal divisor.

Definition 1.2.26. A divisor $D \in \mathrm{Div}(E)$ is principal if it has the form $D = \mathrm{div}(f)$ for some $f \in \overline{k}(E)^*$. Two divisors are linearly equivalent, denoted by $D_1 \sim D_2$, if $D_1 - D_2$ is principal.

Pairings

Two of the most used pairings in cryptography are the Weil pairing and the Tate pairing, both used in subsequent chapters.

Let E/k be an elliptic curve, $P \in E[N]$ a point on E, and $f_{N,P} \in \overline{k}(E)$ a Miller function i.e. a function on E with the divisor equal to $N(P) - N(\mathcal{O}_E)$. There exists a function $g_{N,P} \in \overline{k}(E)$ such that

$$f_{N,P} \circ [N] = g_{N,P}^N. \tag{1.5}$$

Furthermore, their divisors are Galois invariant and [64, Chapter II, Proposition 5.8.] implies that we can choose $f_{N,P}$ and $g_{N,P}$ to be in $k(E)$. The function $g_{N,P}$ can be used to define the Weil pairing. Let $S \in E[N]$ be an N-torsion point where we allow $S = P$. For any other point $X \in E$ we have

$$g_{N,P}(X+S)^N = f_{N,P}([N]X + [N]S) = f_{N,P}([N]X) = g_{N,P}(X)^N.$$

The function $\dfrac{g_{N,P}(X+S)}{g_{N,P}(X)}$ is, for every X, a N^{th} root of unity, i.e. it takes only finitely many values and we can define a pairing

$$e_N \colon E[N] \times E[N] \to \mu_N$$

with

$$e_N(S,P) = \frac{g_{N,P}(X+S)}{g_{N,P}(X)},$$

where $X \in E$ is any point such that $g_{N,P}(X+S)$ and $g_{N,P}(X)$ are both defined and nonzero. The following proposition gives the basic properties of the Weil pairing:

Proposition 1.2.27. The Weil pairing e_N has the following properties:

(i) It is bilinear:

$$e_N(S_1 + S_2, P) = e_N(S_1, P) e_N(S_2, P),$$
$$e_N(S, P_1 + P_2) = e_N(S, P_1) e_N(S, P_2).$$

(ii) It is alternating:

$$e_N(P,P) = 1.$$

(iii) It is nondegenerate:

$$\text{if } e_N(S,P) = 1, \quad \text{for all} \quad S \in E[N], \quad \text{then } P = \mathscr{O}_E.$$

(iv) It is Galois invariant:

$$e_N(S,P)^\sigma = e_N(S^\sigma, P^\sigma), \quad \forall \sigma \in G_{\bar{k}/k}.$$

(v) It is compatible:

$$e_{N,N'}(S,P) = e_N([N']S,P), \quad \text{for all} \quad S \in E[NN'] \text{ and } P \in E[N].$$

Proof. See [64, Chapter III, Proposition 8.1.] ■

From the alternating property we have $e_N(S,P) = e_N(P,S)^{-1}$. An alternative to the Weil pairing is the Tate pairing (or Tate-Lichtenbaum pairing). The definition of this pairing over a field k is available in [64, Chapter XI, Section 9.]. The Tate pairing is a well-defined bilinear pairing, see [64, Chapter XI, Proposition 9.1.]

There is a connection between divisors, the Miller function, and the Tate pairing. Let k be a field, and N an integer such that $\text{char}(k) \nmid N$. Let E/k be an elliptic curve, $P_1 \in E(k)[N]$ and $P_2 \in E(k)/NE(k)$. A Miller function f_{N,P_1} is any function on E with divisor $N(P_1) - N(\mathscr{O}_E)$. To calculate the Tate pairing, i.e. a bilinear map

$$t_N \colon E(k)[N] \times E(k)/NE(k) \to k^*/(k^*)^N \colon (P_1,P_2) \mapsto t_N(P_1,P_2),$$

where $E(k)[N]$ consists of all the points in $E[N]$ defined over k, we let D be a k-rational divisor on E that is linearly equivalent to $(P_2) - (\mathscr{O}_E)$, and the support of which is disjoint from $\{P_1, \mathscr{O}_E\}$.[2] The support of this divisor is disjoint from the divisor of the Miller function f_{N,P_1}, thus

$$f_{N,P_1}(D) = \prod_{P \in E} f_{N,P_1}(P)^{n_P}$$

[2] As an example, we can take $D = (P_2) - (\mathscr{O}_E) + \text{div}(f)$, where $\text{div}(f) = \mathscr{O}_E - (P_2) + (P_2 + T) - (T)$, and $T \notin \{-P_2, \mathscr{O}_E, P_1, P_1 - P_2\}$ is an arbitrary point.

is well-defined. Then, the Tate pairing can be calculated as

$$t_N(P_1,P_2) = f_{N,P_1}(D).$$

Furthermore, if $P_1 \neq P_2$ and the Miller function is normalized[3], the Tate pairing $t_N(P_1,P_2)$ is equal to $f_{N,P_1}(P_2)$.

We refer to [64, Chapter III, Section 8.] and [69, Chapter 3.] for more details on pairings.

1.2.3. Elliptic curves over finite fields

Elliptic curves over finite fields are important for applications in cryptography and the factorization of large integers. In this section, we will review some of the basic properties of such curves, mostly related to the number of rational points and the structure of the group $E(k)$.

Definition 1.2.28. Let $p > 3$ be a prime. An elliptic curve over $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$ is given by an equation of the form

$$E\colon y^2 = x^3 + ax + b, \quad \text{with } a,b \in \mathbb{F}_p, \text{ such that } 4a^3 + 27b^2 \neq 0. \tag{1.6}$$

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

$$E(\mathbb{F}_p) = \{(x,y) \in \mathbb{F}_p \times \mathbb{F}_p\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6\} \cup \{\mathcal{O}_E\}.$$

Remark. Addition formulas from Theorem 1.2.4 applied to the points of elliptic curve (1.6) give a point in $E(\mathbb{F}_p)$. The composition law from Definition 1.2.2 makes $E(\mathbb{F}_p)$ into a finite group.

If we want to estimate the number of points in $E(\mathbb{F}_p)$ we could try to find all solutions to the equation (1.6). For a small p we would calculate y for every value of x. An obvious upper bound is $2p+1$, since for every value of x we get two values for y. A more rigorous bound is given in the following theorem.

[3] The leading coefficient of the function is equal to 1 when expanded in terms of the uniformizer x/y at $\mathcal{O}_E$.

Theorem 1.2.29 (Hasse)**.** Let $E/\mathbb{F}_p$ be an elliptic curve defined over a finite field. Then

$$|\#E(\mathbb{F}_p) - p - 1| \leq 2\sqrt{p}.$$

Proof. See [64, Chapter V, Theorem 1.1.]. ■

We can also go in the other direction: if m is an integer such that

$$m \in \langle p+1-2\sqrt{p}, p+1+2\sqrt{p} \rangle,$$

then there is an elliptic curve over $\mathbb{F}_p$ such that $|E(\mathbb{F}_p)| = m$, see [26, Chapter 4, Theorem 14.18.]. The following theorem expounds on the structure of the group $E(\mathbb{F}_p)$.

Theorem 1.2.30. Let $E/\mathbb{F}_p$ be an elliptic curve defined over a finite field. Then

$$E(\mathbb{F}_p) \simeq \mathbb{Z}/n\mathbb{Z} \quad \text{or} \quad E(\mathbb{F}_p) \simeq \mathbb{Z}/n_1\mathbb{Z} \oplus \mathbb{Z}/n_2\mathbb{Z}$$

for some integer $n \geq 1$ or for some integers $n_1, n_2 \geq 1$ with n_1 dividing n_2.

Proof. See [69, Chapter 4, Theorem 4.1.]. ■

For curves over finite fields, there is always a special endomorphism.

Definition 1.2.31. Let E be an elliptic curve defined over a finite field with q elements. The Frobenius endomorphism $\pi\colon E \to E$ is defined as the map

$$\pi\colon (x,y) \mapsto (x^q, y^q).$$

The proof of the Hasse theorem in [64, Chapter V, Theorem 1.1.] uses the Frobenious endomorphism. There are two interesting facts to note:

- $\ker(\pi) = \mathcal{O}_E$,
- $\ker(\pi - 1) = E(\mathbb{F}_q)$.

Chapter 2 will provide examples illustrating the use of finite fields in cryptography.

1.3. MODULAR CURVES

This section provides some basic definitions and results regarding the theory of congruence subgroups, modular curves, and enhanced elliptic curves.

1.3.1. Congruence subgroups

We start with the definition of a modular group.

Definition 1.3.1. The modular group is a group of 2×2 matrices with integer entries and a determinant equal to 1,

$$\mathrm{SL}_2(\mathbb{Z}) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) : a,b,c,d \in \mathbb{Z}, ad-bc = 1 \right\}.$$

Similarly, we define the principal congruence subgroup.

Definition 1.3.2. The principal congruence subgroup of level $N > 0$ is

$$\Gamma(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \colon \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \pmod{N} \right\}.$$

In particular $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$. The reduction modulo N morphism $\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$ induces a homomorphism $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ with kernel $\Gamma(N)$, thus $\Gamma(N)$ is a normal subgroup in $\mathrm{SL}_2(\mathbb{Z})$ of finite index. This homomorphism is a surjection, so there is an induced isomorphism

$$\mathrm{SL}_2(\mathbb{Z})/\Gamma(N) \xrightarrow{\sim} \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}).$$

The index of the subgroup $\Gamma(N)$ in $\mathrm{SL}_2(\mathbb{Z})$ is equal to

$$[\mathrm{SL}_2(\mathbb{Z}) \colon \Gamma(N)] = N^3 \prod_{p \mid N} \left(1 - \frac{1}{p^2}\right),$$

where the product is taken over all primes p dividing N.

Definition 1.3.3. A subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ is a congruence subgroup if $\Gamma(N) \subset \Gamma$ for some $N \in \mathbb{N}$. In this case, Γ is a congruence subgroup of level N.

From this definition, it is obvious that every congruence subgroup Γ has a finite index in $\mathrm{SL}_2(\mathbb{Z})$. Other standard congruence subgroups are

$$\Gamma_1(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \colon \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right) \pmod{N} \right\},$$

$$\Gamma_0(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \colon \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right) \pmod{N} \right\}.$$

These subgroups satisfy

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \mathrm{SL}_2(\mathbb{Z}).$$

Additionally, $\Gamma(N)$ is a normal subgroup of $\Gamma_1(N)$ because the map

$$\Gamma_1(N) \to \mathbb{Z}/N\mathbb{Z}, \quad \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mapsto b \pmod{N}$$

is a surjection with kernel $\Gamma(N)$. This induces an isomorphism

$$\Gamma_1(N)/\Gamma(N) \xrightarrow{\sim} \mathbb{Z}/N\mathbb{Z},$$

so the index $[\Gamma_1(N)\colon \Gamma(N)]$ is equal to N. Similarly, the map

$$\Gamma_0(N) \to (\mathbb{Z}/N\mathbb{Z})^\times, \quad \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mapsto d \pmod{N}$$

is a surjection with the kernel equal to $\Gamma_1(N)$, so $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$ and, using the induced isomorphism

$$\Gamma_0(N)/\Gamma_1(N) \xrightarrow{\sim} (\mathbb{Z}/N\mathbb{Z})^\times,$$

the index $[\Gamma_0(N)\colon \Gamma_1(N)]$ is equal to $\phi(N)$.[4]

1.3.2. Modular and enhanced elliptic curves

Let $\mathscr{H}$ denote the upper half-plane

$$\mathscr{H} = \{\tau \in \mathbb{C}\colon \mathrm{Im}(\tau) > 0\}.$$

Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ and $\tau \in \mathscr{H}$. The action of the modular group $\mathrm{SL}_2(\mathbb{Z})$ on the upper half-plane, also called fractional linear transformation, is defined as

$$\gamma(\tau) = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)(\tau) = \frac{a\tau + b}{c\tau + d}.$$

This is a well-defined action because of the formula

$$\mathrm{Im}(\gamma(\tau)) = \frac{\mathrm{Im}(\tau)}{|c\tau + d|^2}.$$

We are now ready to define a modular curve.

[4] $\phi(N)$ is the Euler totient function, which counts the number of elements from $\{1,\ldots,N\}$ relatively prime to N.

Definition 1.3.4. Let Γ be a congruence subgroup of level N in $\mathrm{SL}_2(\mathbb{Z})$. The modular curve $Y(N)$ is defined as a quotient (set of orbits)

$$Y(N) = \Gamma/\mathscr{H} = \{\Gamma\tau \colon \tau \in \mathscr{H}\}.$$

For the usual congruence subgroups $\Gamma(N), \Gamma_1(N), \Gamma_0(N)$, we have

$$Y(N) = \Gamma(N)/\mathscr{H},$$
$$Y_1(N) = \Gamma_1(N)/\mathscr{H},$$
$$Y_0(N) = \Gamma_0(N)/\mathscr{H}.$$

If the action is extended to $\mathscr{H}^* = \mathscr{H} \cup \mathbb{Q} \cup \{\infty\}$, we can define the following modular curves

$$X(\Gamma) = \Gamma/\mathscr{H}^*,$$
$$X(N) = \Gamma(N)/\mathscr{H}^*,$$
$$X_1(N) = \Gamma_1(N)/\mathscr{H}^*,$$
$$X_0(N) = \Gamma_0(N)/\mathscr{H}^*.$$

Complex elliptic curves

The quotients of the upper half-plane by the congruence subgroups can be described by the sets of equivalence classes of elliptic curves that are enhanced with some torsion data. Elliptic curves enhanced with torsion data are called enhanced elliptic curves.

Definition 1.3.5. Let E be an elliptic curve over an algebraically closed field[5] the characteristic of which does not divide N.

(a) An enhanced elliptic curve for $\Gamma_0(N)$ is an ordered pair (E,C), where C is a cyclic subgroup of E of order N. Two enhanced elliptic curves (E,C) and (E',C') are equivalent if there exists an isomorphism $E \xrightarrow{\sim} E'$ that takes C to C'. We denote the set of equivalence classes of enhanced elliptic curves for $\Gamma_0(N)$ by

$$S_0(N) = \{\text{enhanced elliptic curves for } \Gamma_0(N)\}/\sim .$$

[5] A field k is algebraically closed if every nonconstant polynomial in $k[x]$ has a root in k.

(b) An enhanced elliptic curve for $\Gamma_1(N)$ is a pair (E,P), where P is a point of order N. Two enhanced elliptic curves (E,P) and (E',P') are equivalent if there exists an isomorphism $E \xrightarrow{\sim} E'$ that takes P to P'. We denote the set of equivalence classes of enhanced elliptic curves for $\Gamma_1(N)$ by

$$S_1(N) = \{\text{enhanced elliptic curves for } \Gamma_1(N)\}/\sim .$$

Complex elliptic curves are defined using lattices.

Definition 1.3.6. A lattice in $\mathbb{C}$ is a set $\Lambda = \omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$ with $\{\omega_1,\omega_2\}$ a basis for $\mathbb{C}$ over $\mathbb{R}$. The usual normalizing convention is $\frac{\omega_1}{\omega_2} \in \mathscr{H}$.

Two lattices $\Lambda = \omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$ and $\Lambda' = \omega_1'\mathbb{Z} \oplus \omega_2'\mathbb{Z}$, with $\frac{\omega_1}{\omega_2} \in \mathscr{H}$ and $\frac{\omega_1'}{\omega_2'} \in \mathscr{H}$, are equal if there is a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ such that

$$\begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}.$$

Two lattices Λ, Λ' are homothetic if there exists a complex number τ such that $\Lambda = \tau\Lambda'$.

Definition 1.3.7. Let $\Lambda = \omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$ be a lattice. Let $\tau = \frac{\omega_1}{\omega_2}$ and let $\Lambda_\tau = \mathbb{Z} \oplus \tau\mathbb{Z}$. A complex torus is a quotient of the complex plane by the lattice:

$$\mathbb{C}/\Lambda_\tau = \{z + \Lambda_\tau \colon z \in \mathbb{C}\}.$$

There is a correspondence between elliptic curves and complex tori, i.e. for any homotety class of complex tori there is an isomorphism class of elliptic curves and vice versa, so we denote the quotient $\mathbb{C}/\Lambda_\tau$ by E_τ. For more details about this correspondence, we refer to [64, Chapter 6.].

Example 1.3.8. Let N be a positive integer and Λ a lattice. Multiplication by N is the map

$$[N] \colon \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda, \quad z + \Lambda \mapsto Nz + \Lambda.$$

This is a well-defined map since $N\Lambda \subset \Lambda$. Points $z + \Lambda \in \mathbb{C}/\Lambda$ such that $[N](z+\Lambda) = 0$ are the N-torsion points of $\mathbb{C}/\Lambda$. Because of an elliptic curve - complex tori correspondence, if we let E denote the elliptic curve for complex tori $\mathbb{C}/\Lambda$, N-torsion points are denoted by $E[N]$ and the map $[N]$ is an isogeny.

The definition of the sets $S_0(N)$ and $S_1(N)$ from Definition 1.3.5 remains unchanged when the underlying field is $\mathbb{C}$ and E is a complex elliptic curve. Points of $Y_1(N)$ are in bijection with isomorphism classes of pairs $(E,P) \in S_1(N)$. To establish this bijection, to $\tau \in \mathscr{H}$, associate the pair $(E_\tau, \frac{1}{N} + \Lambda_\tau)$. Any pair (E,P) is isomorphic to $(E_\tau, \frac{1}{N} + \Lambda_\tau)$ for some $\tau \in \mathscr{H}$ and E_τ is isomorphic to $E_{\tau'}$ if and only if $\tau' \in \Gamma_1(N)\tau$. So, we have the following theorem:

Theorem 1.3.9. Let N be a positive integer. The moduli space for $\Gamma_1(N)$ is

$$S_1(N) = \left\{ \left[E_\tau, \frac{1}{N} + \Lambda_\tau \right] : \tau \in \mathscr{H} \right\}.$$

Two points $[E_\tau, \frac{1}{N} + \Lambda_\tau]$ and $[E_{\tau'}, \frac{1}{N} + \Lambda_{\tau'}]$ are equal if and only if $\Gamma_1(N)\tau = \Gamma_1(N)\tau'$. Thus, there is a bijection

$$\psi_1 : S_1(N) \xrightarrow{\sim} Y_1(N), \quad [\mathbb{C}/\Lambda_\tau, \frac{1}{N} + \Lambda_\tau] \mapsto \Gamma_1(N)\tau.$$

Proof. See [31, Chapter 1, Theorem 1.5.1.]. ■

Theorem 1.3.9 has analogous versions for congruence subgroups $\Gamma_0(N)$ and $\Gamma(N)$, also part of [31, Chapter 1, Theorem 1.5.1.].

An alternative definition of modular curves

The definition of modular curves from the previous section is a standard (or classical) one. Starting from some congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$, the quotient of the upper half-plane by Γ is the modular curve $Y(\Gamma)$, and the quotient of the extended upper half-plane by the same congruence subgroup (i.e. adding the cusps) is the modular curve $X(\Gamma)$. Moreover, as we have seen in Theorem 1.3.9, equivalence classes of elliptic curves with torsion data are parametrized with modular curves constructed from the usual congruence subgroups $\Gamma(N)$, $\Gamma_1(N)$ and $\Gamma_0(N)$.

Let N be a positive integer. In this subsection, to determine a field of definition of automorphisms of a modular curve in Section 4.2.2, we will generalise the above-mentioned construction for an arbitrary subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$, where $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ is the group of invertible 2×2 matrices with entries from $\mathbb{Z}/N\mathbb{Z}$. We are following [63]. Let H be a

subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$. To the subgroup H associate the congruence subgroup $\Gamma_H(N)$ defined as

$$\Gamma_H(N) \coloneqq \{M \in \mathrm{SL}_2(\mathbb{Z}) \colon (M \bmod N) \in H \cap \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})\}. \tag{1.7}$$

In other words, the congruence subgroup $\Gamma_H(N)$ is the preimage of $H \cap \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ under the reduction modulo N map $\pi_N \colon \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. Moreover, $\Gamma_H(N)$ is a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, as $\Gamma(N) = \{M \in \mathrm{SL}_2(\mathbb{Z}) \colon M \equiv I \pmod N\} \subseteq \Gamma_H(N)$. Over the complex numbers, the quotient $Y(\Gamma_H(N)) \coloneqq \Gamma_H(N)/\mathscr{H}$ can be compactified to the modular curve $X(\Gamma_H(N)) \coloneqq \Gamma_H(N)/\mathscr{H}^*$. This curve has a model defined over $\mathbb{Q}(\zeta_N)^{\det(H)}$, and when the map $\det \colon H \to (\mathbb{Z}/N\mathbb{Z})^\times$ is surjective, it has a model defined over $\mathbb{Q}$.

The modular curve $X(\Gamma_H(N))$, similar to the classical case, parametrizes equivalence classes of elliptic curves. Those equivalence classes are dependent on the subgroup H.

Definition 1.3.10. Let E be an elliptic curve over $\mathbb{C}$, and let N be a positive integer. A level N-structure is defined as an isomorphism $\phi \colon (\mathbb{Z}/N\mathbb{Z})^2 \to E[N]$.

Remark. A level N-structure is a choice of basis for $E[N]$. It could be also defined as an isomorphism $E[N] \to (\mathbb{Z}/N\mathbb{Z})^2$.

Let (E,ϕ) and (E',ϕ') be pairs of an elliptic curve and a level N-structure. For the moduli interpretation, we need an equivalence relation between such pairs together with their connection to the subgroup H. In the following definition, an element $h \in H$ is considered as an isomorphism $h \colon (\mathbb{Z}/N\mathbb{Z})^2 \to (\mathbb{Z}/N\mathbb{Z})^2$.

Definition 1.3.11. Pairs (E,ϕ) and (E',ϕ') are H-isomorphic if there is an isomorphism $\iota \colon E \to E'$ and an element $h \in H$ such that:

$$h = (\phi')^{-1} \circ \iota|_{E[N]} \circ \phi.$$

Being H-isomorphic is an equivalence relation, denoted by $(E,\phi) \sim_H (E',\phi')$. The equivalence class is denoted by $[(E,\phi)]_H$.

To complete the moduli interpretation, let $E_\tau = \mathbb{C}/\Lambda_\tau$ be a complex elliptic curve, for $\tau \in \mathscr{H}$ and $\Lambda_\tau = \mathbb{Z} \oplus \tau\mathbb{Z}$. For the usual choice of basis for $E_\tau[N]$, let the level N-structure

map $\phi_\tau\colon (\mathbb{Z}/N\mathbb{Z})^2 \to E_\tau[N]$ be the map $(1,0)\mapsto 1/N$ and $(0,1)\mapsto \tau/N$. We have the following bijection:

$$\begin{aligned}\{[(E_\tau,\phi_\tau)]_H\colon \tau\in\mathscr{H}\} &\to Y(\Gamma_H(N))\\ [(E_\tau,\phi_\tau)]_H &\mapsto \Gamma_H(N)\tau.\end{aligned} \tag{1.8}$$

The existence of the bijection (1.8) can be proven similarly to the proof of Theorem 1.3.9. In other words, $Y(\Gamma_H(N))$ parametrizes elliptic curves with the added H-structure.

The modular interpretation (1.8) can be used to define modular curves over more general objects, by utilizing elliptic curves over arbitrary base schemes. The following is a quick summary using [34] and [14], and the usual references for more details are [42] or [40]. We will introduce some background and results regarding the theory of schemes with most of the details omitted. Informally, we will consider a scheme S to be a more general object than an algebraic variety. An elliptic curve over S is a pair $(E\to S,\mathscr{O}_E)$, where the map $E\to S$ is proper and smooth, all fibres are geometrically connected curves of genus 1, and $\mathscr{O}_E$ is a (zero) section of $E\to S$. The elliptic curve over a scheme is denoted, as usual, by E/S, and it has the structure of a commutative group scheme.

Let (E,ϕ) denote a pair of an elliptic curve E over a $\mathbb{Q}$-scheme S, and $\phi\colon (\mathbb{Z}/N\mathbb{Z})^2_S \to E[N]$ an isomorphism of S-group schemes. This is an analogue to the level N-structure from before, called a full-level N-structure, and a modular curve that parametrizes pairs (E,ϕ) is denoted by $Y(N)$. Its compactification is denoted by $X(N)$ and called the modular curve of full-level N.

Let $\gamma\in \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ be a matrix. Every such matrix gives an automorphism of the constant group scheme $(\mathbb{Z}/N\mathbb{Z})^2_S$, so γ acts on $Y(N)$ by sending pair (E,ϕ) to pair $(E,\phi\circ\gamma)$. This defines an action of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ on $Y(N)$, which can be uniquely extended to $X(N)$. Then, for every subgroup H of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ we let X_H be the quotient $X(N)/H$. The model of $Y_H = Y(N)/H$ over $\mathbb{Z}[1/N]$ is a coarse moduli space for elliptic curves with H-structure, i.e. for pairs of elliptic curve E over a $\mathbb{Z}[1/N]$-scheme S, where level N-structure is the map $\phi\colon (\mathbb{Z}/N\mathbb{Z})^2_S\to E[N]$, an isomorphism of S-group schemes, and the equivalence relation is given by (1.8). This is also valid for an algebraically closed field k, whose characteristic does not divide N, i.e. we have a bijection between $Y_H(k)$ and elliptic curves over k with H-structure.

Remark. At this point, for a subgroup $H\in \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$, we have two (definitions

of) modular curves, one is $X(\Gamma_H(N))$, and the other one is X_H. Those modular curves are of course connected. Over the complex numbers, if $\det(H) = (\mathbb{Z}/N\mathbb{Z})^\times$, the curve $X_H(\mathbb{C})$ is connected, and equal to $X(\Gamma_H(N))$. If $\det(H) \neq (\mathbb{Z}/N\mathbb{Z})^\times$, then the curve is not connected, and the number of connected components of $X(\Gamma_H(N))$ is equal to $[(\mathbb{Z}/N\mathbb{Z})^\times : \det(H)]$.

The modular curve X_H will be used in Section 4.2 to determine the field of definition for automorphisms of modular curves.

2. ISOGENY-BASED POST-QUANTUM CRYPTOGRAPHY

In this chapter, we will cover the basics of cryptography and post-quantum cryptography, and give an overview of isogeny-based post-quantum cryptography. We will mostly follow [65] and [70].

2.1. INTRODUCTION TO CRYPTOGRAPHY

The primary purpose of cryptography is to ensure the secure transmission of sensitive data or messages through an insecure channel. This is achieved by allowing only the sender (often referred to as Alice) and the intended recipient (usually referred to as Bob) the possibility of understanding the message. At its core, cryptography involves working with the original message, known as plaintext, along with a cryptographic key and an algorithm or scheme that utilizes the plaintext and the key to generate an unreadable message, known as ciphertext. A collection of three algorithms, one for key generation, one for encryption, and one for decryption is called a cryptosystem.

Depending on the type of key used, cryptosystems can be either symmetric or asymmetric. In symmetric cryptography, the sender and the recipient share the same key. The sender uses the key to encrypt the message, and the recipient to decrypt it. The main disadvantage of this principle is that they need to exchange the symmetric key before secure communication can start. The symmetric cryptosystem is visually represented in Figure 2.1. One of the most known and used symmetric cryptosystems is the Advanced Encryption Standard (AES), which became the NIST standard in 2006, see [36] for the description of the algorithm.

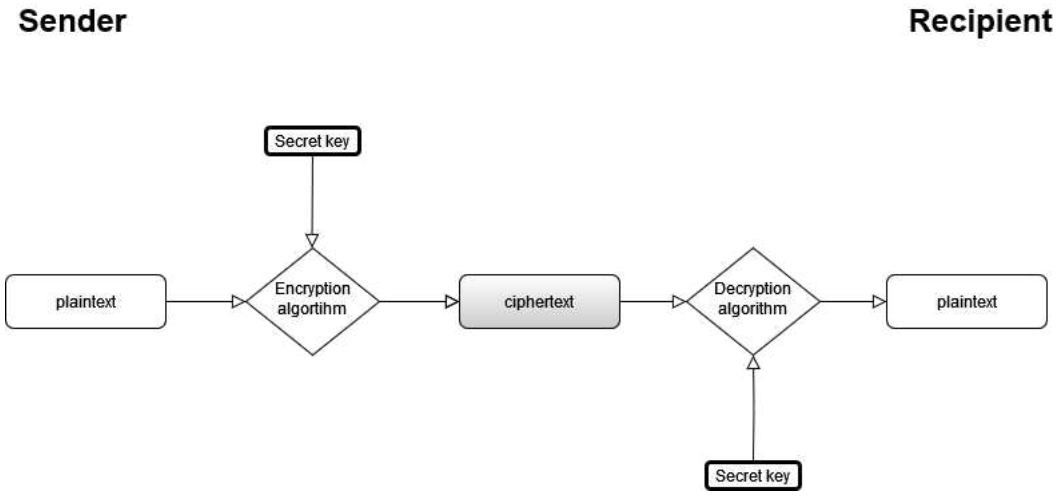

Figure 2.1: Symmetric encryption

Symmetric cryptography remained the sole method used until 1976, and the introduction of asymmetric cryptography in [32]. In this type of cryptography, all participants have a pair of keys: one secret and used for decryption, and the other public (asymmetric cryptography is also called public-key cryptography) and used for encryption. Asymmetric cryptography's primary advantage is that there is no need to exchange keys before secure communication can begin. The security of public-key cryptosystems relies on the existence of a trapdoor one-way function, i.e. an invertible function that is easy to calculate in one direction but difficult to compute in reverse, except when the auxiliary trapdoor information is known. An asymmetric cryptosystem is visually represented in Figure 2.2.

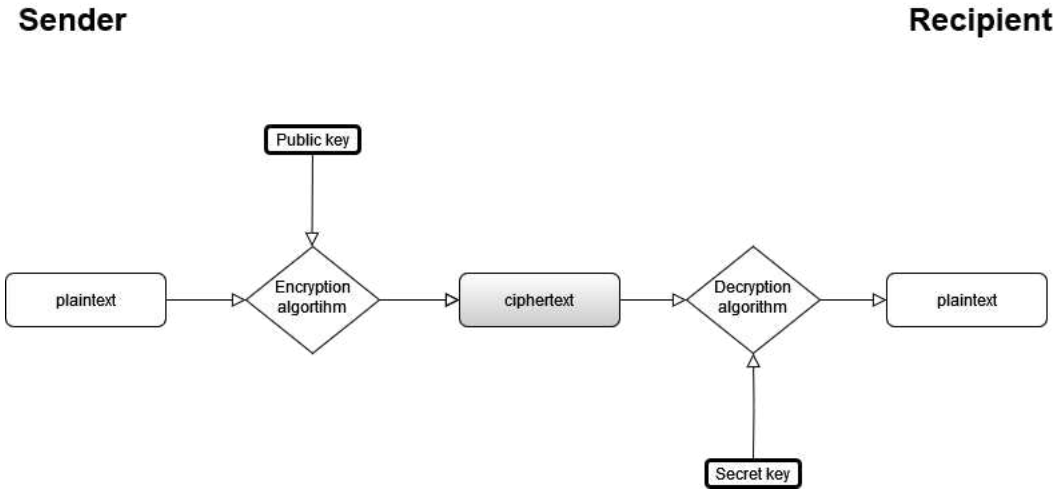

Figure 2.2: Asymmetric encryption

In the next section, we will introduce some basic ideas concerning public-key cryptography, as this branch of cryptography has been particularly affected by the development of quantum computers.

2.1.1. Public-key cryptography

The discrete logarithm problem is the basis for a number of public-key cryptographic constructions. We will start with the most general definition of the problem in a group setting.

Definition 2.1.1. Let $(G,\cdot)$ be a finite group. The discrete logarithm problem (DLP) for G is to determine, for an element $g \in G$, and an element $h \in \langle g \rangle$, the least positive integer x satisfying

$$\underbrace{g \cdot g \cdots g \cdot g}_{x \text{ times}} = h.$$

One of the first examples of DLP is the previously mentioned work introduced in [32], known as the Diffie-Hellman key exchange, a way to exchange a shared secret (usually a symmetric key) between the sender and the recipient. The description of the Diffie-Hellman key exchange, when the underlying group is the finite field $\mathbb{F}_p^*$, can be summarized as:

- Alice and Bob agree on a large prime p and a nonzero integer g having a large order in $\mathbb{F}_p^*$. Those values are public.
- Alice picks a secret integer a, while Bob picks a secret integer b. Those integers are used to compute values

$$A \equiv g^a \pmod{p} \quad \text{and} \quad B \equiv g^b \pmod{p}.$$

- The public values A and B are exchanged.
- Using their secret integers, Alice can compute

$$A' \equiv B^a \pmod{p},$$

  and Bob can compute

$$B' \equiv A^b \pmod{p}.$$

- The values A' and B' are the same since

$$A' \equiv B^a \equiv (g^b)^a \equiv g^{ab} \equiv (g^a)^b \equiv A^b \equiv B' \pmod{p}.$$

- This common value is their exchanged key, i.e. their shared secret.

The algorithm that efficiently computes large powers of number g modulo number p is called the Square-and-multiply algorithm, see [65, Section 1.3.2.]. An example of a public-key cryptosystem based on a discrete algorithm problem, and a direct extension of the Diffie-Hellman key exchange is Elgamal, see [65, Section 2.4.].

Elliptic curve discrete logarithm problem

Given the definition of the discrete logarithm problem and the existence of a group law on the points of an elliptic curve, it is natural to assume that the discrete logarithm problem can be extended to the elliptic curve setting.

Let E be an elliptic curve over a finite field $\mathbb{F}_p$. Alice chooses a point P, an integer n and computes $Q = [n]P$. She publishes points P and Q and n remains secret. An attacker (usually referred to as Eve) needs to find out how many times should she add P to get Q.

Definition 2.1.2. Let E be an elliptic curve over the field $\mathbb{F}_p$, and let P and Q be points in $E(\mathbb{F}_p)$. The elliptic curve discrete logarithm problem (ECDLP) is the problem of finding an integer n such that $Q = [n]P$.

The Diffie-Hellman key exchange for elliptic curves is:

- Alice and Bob agree on a large prime p, an elliptic curve E over $\mathbb{F}_p$, and a point P in $E(\mathbb{F}_p)$. Those parameters are public.

- Alice chooses a secret integer n_A and computes the point

$$Q_A = [n_A]P.$$

  Bob chooses a secret integer n_B and computes the point

$$Q_B = [n_B]P.$$

- They exchange Q_A and Q_B.

- Alice computes the point $[n_A]Q_B$ and Bob computes the point $[n_B]Q_A$.
- The shared secret is

$$[n_A]Q_B = [n_A]([n_B]P) = [n_B]([n_A]P) = [n_B]Q_A.$$

An attacker Eve needs to solve ECDLP of the form $[n_A]P = Q_A$ to get the shared secret used between Alice and Bob. Similarly to the Square-and-multiply algorithm mentioned before, the algorithm that efficiently computes $[n_A]P, [n_B]P$ is called the Double-and-add algorithm, see [65, Section 6.3.1.].

The concept of using elliptic curves in cryptography was first independently proposed by Koblitz [45] and Miller [51]. They suggested that the ECDLP might be more difficult than the DLP in a finite field. Consequently, elliptic curve cryptosystems should require smaller key sizes while offering better performance. Despite these advantages, it was not until the late 2000s that elliptic curves became widely used and started to prevail over other public-key cryptosystems.

An example of a public-key cryptosystem based on an elliptic curve discrete algorithm problem, and a direct extension of the Elgamal algorithm to elliptic curves, is the Menezes-Vanstone algorithm, see [65, Chapter 6, Table 6.13.].

RSA algorithm

In 1977 in [59], Rivest, Shamir, and Adleman (hence the name RSA), introduced one of the first public-key cryptosystems, the security of which lies in the difficulty of factoring large numbers. The basic description of the algorithm is as follows:

- **Key creation**:

  Alice chooses secret primes p and q. Alice also chooses encryption exponent e with

$$\gcd(e, (p-1)(q-1)) = 1.$$

  The modulus $N = pq$ and the exponent e are public.

- **Encryption**:

  Bob chooses the plaintext m and uses Alice's public key (N, e) to compute

$$c \equiv m^e \pmod{N}.$$

He sends the ciphertext c to Alice.

- **Decryption**:
  Alice computes d satisfying

$$ed \equiv 1 \pmod{(p-1)(q-1)},$$

  and then computes

$$m' \equiv c^d \pmod{N}.$$

  The computed m' is equal to the plaintext m.

The assumption behind RSA is that solving $x^e \equiv c \pmod{N}$ is easy for the person who knows the values p and q, because they can be used to calculate value $\phi(N) = (p-1)(q-1)$ and consequently d, but hard for anybody else (an attacker). In the secure implementations of RSA, the modulus N should be several hundred digits long.

Digital signatures

Digital signatures are a branch of public-key cryptography designed to replace physical signatures on paper. In this scenario, Alice, as an owner of a digital document, creates additional data directly linked to the document and uses it to prove that the digital document is uniquely associated with her. A digital signature scheme consists of two main components, the signing algorithm and the verification algorithm. In its simplest form, it can be summarized as:

- Let D denote the document that Alice wants to sign. The private and the public key are denoted by K_{pri} and K_{pub}.
- Alice uses a signing algorithm that takes her private key K_{pri} and the document D as inputs and returns signature D_{sign}.
- Bob uses a verification algorithm that takes document D, signature D_{sign}, and the public key K_{pub} as inputs, and the output returns true if the verification is successful and false if not.

Any signature scheme should at least satisfy:

- From the public key K_{pub}, an attacker cannot determine the private key K_{pri}, or a different key that can produce the same signature D_{sign}.
- From the public key K_{pub}, the list of documents $D^1, \ldots, D^n$ and their signatures $D^1_{sign}, \ldots, D^n_{sign}$, an attacker cannot determine a valid signature for any other new document. In other words, the knowledge of any number of document-signature pairs does not provide the attacker new, useful information.

The following is the RSA digital signature variant.

- **Key creation**:
  Alice chooses secret primes p and q. Alice also chooses verification exponent e with
  $$\gcd(e, (p-1)(q-1)) = 1.$$
  Values $N = pq$ and e are public.
- **Signing**:
  Alice computes d satisfying
  $$de \equiv 1 \pmod{(p-1)(q-1)}.$$
  Alice signs document D by computing
  $$S \equiv D^d \pmod{N}.$$
- **Verification**:
  Bob computes $S^e \pmod{N}$ and verifies that is equal to D.

Further examples of digital signature schemes are given in Section 2.2.2.

2.2. INTRODUCTION TO POST-QUANTUM CRYPTOGRAPHY

Post-quantum cryptography is an area of cryptography that focuses on developing cryptosystems resistant to attacks from both classical and quantum computers. We will start with a short introduction to quantum computing and quantum computers. More details can be found in [70, Chapter 14.].

2.2.1. Quantum computing

When observed on a small scale, particles, such as atoms and subatomic particles tend to behave differently than what is expected in classical physics. For a brief moment, particles can behave like waves, where different waves can superpose to merge into bigger waves, or they can cancel each other out. This behavior is studied by quantum physics (mechanics).

The development of quantum computers is based on two quantum physics phenomena. The first is the quantum superposition, a way for a particle or an object to exist in more than one state simultaneously. An object will remain in this state until measured, after which it collapses in one of the states, each one with a certain probability of being observed. Quantum superposition is the basis for a qubit, the quantum analogue of a classical bit. Qubit is a bit that can be 0 or 1, or both of them at the same time. The second important principle is quantum entanglement. Multiple qubits can become entangled with each other, even when they are not physically connected, and measuring one of the qubits has a direct correlation to the state of the other qubits.

The idea of quantum computing, i.e. building a computer based on the physical phenomena from quantum mechanics, first appeared in 1980. American physicist Benioff [10], was the first to describe a quantum computer, and, in a way, introduced the quantum Turing machine. At the time of writing this thesis, IBM has developed a 433-qubit computer[1].

[1] https://newsroom.ibm.com/2022-11-09-IBM-Unveils-400-Qubit-Plus-Quantum-Processor-and-Next-Generation-IBM-Quantum-System-Two

The first impact of quantum computers on cryptography came in 1994, when Shor [62], demonstrated that there is an efficient quantum algorithm (an algorithm that runs on the quantum computer) for finding discrete logarithms and factoring integers. When/if a big enough quantum computer is built, this algorithm, appropriately called Shor's algorithm, breaks the basic primitives used in asymmetric cryptography. The number of qubits needed for the factorization of the big numbers used by today's modern cryptography is not known, it is of course a lot more than 433, nevertheless, it is evident that quantum-safe protocols need to be developed.

We will briefly mention the impact of quantum computers on symmetric cryptography. In 1996 in [39], Grover introduced a quantum search algorithm, i.e. an algorithm that optimizes a search of an unordered list. For an unordered list of N elements, a search on a classical computer takes on average $N/2$ operations, while the Groves algorithm can perform the same search in $\sqrt{N}$ operation. This algorithm could, for example, be used for brute-forcing a symmetric key. A 128 bit key could be broken on a quantum computer in 2^{64} operations (2^{127} on the classical computer). The "fix" for the Grover algorithm is easier: increasing the length of the symmetric key used to at least 256 bits.

2.2.2. Areas of post-quantum cryptography

Post-quantum cryptography is a direct response of the cryptographic community to the quantum "threat". In an effort to gather and standardize post-quantum cryptography, NIST announced at PQCrypto in 2016 and later issued a formal call[2] for quantum-resistant encryption schemes (or key encapsulation mechanism - KEM) and signature schemes.

The submission deadline for the first round was in November 2017. In the initial round, there were 69 accepted candidates, out of which 5 were later withdrawn. Among the accepted candidates, there were 45 encryption/KEM schemes and 19 signature schemes. A number of submissions (25) were broken during the first round of evaluation, and some of them did not satisfy the NIST security criteria. Consequently, only 17 encryption/KEM schemes and 9 signature schemes were chosen to advance to the second round, see [2].

The second round started in January 2018 and lasted until July 2020. During this

[2] https://csrc.nist.gov/Projects/post-quantum-cryptography/post-quantum-cryptography-standardization/Call-for-Proposals

round, from 26 candidates, 7 schemes were attacked or broken, leaving 15 schemes to proceed to the third round, see [53]. The 15 schemes were divided into two categories: finalists and alternate candidates. The finalists, 4 encryption/KEM schemes and 3 signature schemes, were considered most likely to be ready for standardization soon after the end of the third round. The alternate candidates, 5 encryption/KEMs schemes and 3 signature schemes, were considered ready to be standardized after one more round of evaluation. The third round concluded in July 2022 [3], and as a result, one encryption/KEM scheme and three signature schemes were selected for standardization.

From the NIST standardization process, several bigger areas of post-quantum cryptography became apparent. While many schemes were specifically designed for this standardization effort, some of them have been around for several decades. We give an informal overview of these areas, along with references to more detailed information:

- Lattice-based cryptography ([11, Chapter 5.]):
  Let $v_1, \ldots, v_n \in \mathbb{R}^m$ be a set of linearly independent vectors. A lattice L generated with vectors $v_1, \ldots, v_n$ is the set

  $$L = \{a_1 v_1 + \cdots + a_n v_n,\ a_1, \ldots, a_n \in \mathbb{Z}\}.$$

  The security of lattice-based cryptography is based on the hardness of lattice problems, more notably the shortest vector problem (SVP), and its variants. In an SVP, given a lattice as input, the goal is to return the shortest (exact or approximate) nonzero vector in that lattice.

  This is the most promising area for post-quantum cryptography. Many lattice-based cryptosystems are efficient and easy to implement and the security claim behind them is strong and well-studied. Currently, there are no known quantum algorithms capable of solving lattice problems that perform significantly faster than the best-known classical algorithms.

  It is not surprising that in both categories, KEM and signatures, NIST selected lattice-based cryptosystems. Specifically, for KEM, the selected system is CRYSTALS-KYBER [7], and for signatures, NIST selected CRYSTALS-DILITHIUM [9] and Falcon [37]. Other notable lattice-based cryptosystems are the first one ever, introduced by Ajtai in [1] or NTRU [21], a third-round finalist.

- Multivariate cryptography ([11, Chapter 6.]):
  Multivariate public-key cryptography is based on multivariate polynomials (a polynomial in more than one variable) over a finite field. The public key is given by a set of quadratic polynomials, and the evaluation of these polynomials at any given value represents encryption/decryption. Notable schemes include Oil and Vinegar [44], and until recently Rainbow [33], a third-round finalist broken in [13].

- Code-based cryptography ([11, Chapter 4.]):
  In code-based cryptography, the error-correcting code is an underlying one-way function. The first example of a public-key code-based encryption scheme was introduced by McEliece in [48]. A variant of this scheme [4], called the Classic McEliece, is a fourth-round finalist. The private key is the binary irreducible Goppa code, the public key is the random generator matrix of a randomly permuted version of that code and the ciphertext is a codeword with added errors, removable only by the Goppa code's owner. Code-based cryptography is characterized by the large public key size, but fast encryption and decryption. Other notable examples include BIKE [5] and HQC [49], both of them KEM schemes and alternate candidates in the third and candidates in the fourth round.

- Hash-based cryptography ([11, Chapter 3.]):
  Hash-based schemes are mostly digital signature schemes built upon a cryptographic hash function; their security relies on collision resistance[3] of the hash function. The first hash-based signature scheme was introduced by Merkle in [50]. Another notable scheme is SPHINCS+ [6], selected as one of the signature schemes for standardization.

- Isogeny-based cryptography:
  The following section will delve deeper into isogeny-based cryptography. The only isogeny-based candidate for NIST standardization was SIKE [8], a fourth-round candidate, but the algorithm was completely broken in various papers [15, 47, 60] in 2022.

[3] Hash function is collision-resistant if it is hard to find two different inputs of the same hash value.

2.3. SELECTED SCHEMES FROM ISOGENY-BASED CRYPTOGRAPHY

In the introduction, we mentioned the main developments in the history of isogeny-based cryptography. This includes the initial concept of using isogenies proposed by Couveignes in 1997 (which was also independently published by Rostovtsev and Stolbunov in 2006), the introduction of the first hash function by Charles, Goren, and Lauter in 2006, the key exchange SIDH by de Feo and Jao in 2011, the alternative key exchange approach CSIDH introduced by Castryck, Lange, Martindale, Panny, and Renes in 2018, and the digital signature scheme SQISign introduced by De Feo, Kohel, Leroux, Petit, and Wesolowski in 2020, to mention but a few. The goal of this section is to provide an overview of the CSIDH key exchange.

2.3.1. Isogeny graphs

Isogeny-based cryptography is based on isogeny graphs. This section provides some background on graphs and isogeny graphs, following [27]. We start with the definition of a graph.

Definition 2.3.1. An undirected graph G is a pair (V,E) where V is a finite set of vertices and $E \subset V \times V$ is a set of unordered pairs called edges.

Two vertices v and v' are connected by an edge if $\{v,v'\} \in E$. A path between two vertices v,v' is a sequence of vertices $v \to v_1 \to v_2 \to \cdots \to v'$ such that each vertex is connected to the next by an edge. A graph is considered connected if any two vertices have a path connecting them, otherwise, it is considered disconnected. The diameter of a connected graph is the largest of all distances between its vertices. The degree of a vertex is the number of edges pointing to (or from) it. A graph where every edge has degree k is called k-regular.

Definition 2.3.2. Let K be a field and l a positive integer not divisible by $\mathrm{char}(K)$. The l-isogeny graph G over K is a graph where the nodes are elliptic curves defined over K,

up to K-isomorphism, and edges are l-isogenies between those elliptic curves if such an isogeny exists.

Commonly, because of Proposition 1.2.11, the nodes in the isogeny graph are defined as j-invariants. From Theorem 1.2.19, we know that every isogeny has a dual isogeny of the same degree, so isogeny graphs are considered undirected. The connected component of an l-isogeny graph, when ignoring special nodes where the j-invariant is equal to 0 or 1728, has one of two shapes, it is a volcano or a Pizer graph. For more details see [56].

We continue with general graph theory. The adjacency matrix of a graph $G = (V,E)$, where $V = \{v_1, \ldots, v_n\}$, is the $n \times n$ matrix where the element at (i, j) is 1 if there is an edge between v_i and v_j and 0 otherwise. When the graph is undirected, the adjacency matrix is symmetric and has n real eigenvalues

$$\lambda_1 \geq \cdots \geq \lambda_n.$$

In a k-regular graph, the largest and smallest eigenvalues λ_1 and λ_n satisfy $k = \lambda_1 \geq \lambda_n \geq -k$. A k-regular graph such that $|\lambda_i| \leq \sqrt{k-1}$ for any λ_i except λ_1, is called a Ramanujan graph.

Definition 2.3.3. Let $\varepsilon > 0$ and $k \geq 1$. A k-regular graph is called a (one-sided) ε-expander if

$$\lambda_2 \leq (1-\varepsilon)k,$$

and a two-sided ε-expander if it also satisfies

$$\lambda_n \geq -(1-\varepsilon)k.$$

Informally, an expander graph is an undirected graph where every subset of the vertices has the expanding property, i.e. it is connected with a large number of vertices from its complement.

Let $G_i = (V_i, E_i)$ be a sequence of k-regular graphs with $\#V_i \to \infty$. This sequence is a one-sided (two-sided) expander family if there is an $\varepsilon > 0$ such that G_i is one-sided (two-sided) for all sufficiently large i.

Theorem 2.3.4. Let $k \geq 1$, and let G_i be a sequence of k-regular graphs. Then

$$\max(|\lambda_2|, |\lambda_n|) \geq 2\sqrt{k-1} - o(1), \quad \text{as } n \to \infty.$$

In a graph $G = (V,E)$, a random walk of length i is a path $v_1 \to \cdots \to v_i$, defined by randomly selecting vertices v_j uniformly among the neighbours of v_{j-1}. The mixing property of an expander graph says that the random walks of length close to its diameter terminate on any vertex with probability close to uniform. According to the next theorem, supersingular graphs are Ramanujan.

Theorem 2.3.5. Let p and l be distinct primes.

(i) All supersingular j-invariants of curves in $\overline{\mathbb{F}}_p$ are defined in $\mathbb{F}_{p^2}$.

(ii) The number of isomorphism classes of supersingular elliptic curves over $\overline{\mathbb{F}}_p$ is equal to

$$p + \begin{cases} 0, & \text{if } p \equiv 1 \pmod{12}, \\ 1, & \text{if } p \equiv 5,7 \pmod{12}, \\ 2, & \text{if } p \equiv 11 \pmod{12}. \end{cases}$$

(iii) The graph of supersingular curves in $\overline{\mathbb{F}}_p$ with l-isogenies is connected, $l+1$ regular, and has the Ramanujan property.

Proof. See [27, Part III, Theorem 47.]. ■

Theorem 2.3.5 implies that isogeny graphs have a good mixing property, which makes them suitable for cryptography. For example, the Charles-Goren-Lauter hash function [20] starts from an arbitrary vertex v_1 in an expander graph, then takes a random walk (without backtracking) according to the string to be hashed, and outputs the arrival vertex. Previously mentioned schemes CRS and CSIDH both operate on a union of several large subgraphs on the same vertex set.

2.3.2. Diffie-Hellman in an isogeny setting

In this section, we will provide a brief overview of one of the most known isogeny-based schemes, the CSIDH (commutative-SIDH), keeping with the original paper [18]. The theory behind this scheme is complex and goes beyond the scope of this thesis. However, as we will see in the next chapter, CSIDH benefits from radical isogenies. We start with Couveignes's definition of a hard homogeneous space.

Definition 2.3.6. A hard homogeneous space consists of a finite commutative group G acting freely and transitively on some set S. We denote the action with $*$. The following operations are required to be easy (e.g., polynomial-time):

(i) Compute the group operations in G.

(ii) Sample randomly from G with (close to) uniform distribution.

(iii) Decide validity and equality of a representation of elements of S.

(iv) Compute the action of a group element $g \in G$ on some $s \in S$.

The following problems are required to be hard (e.g., not polynomial-time):

(i) Given s, $s' \in S$, find $g \in G$ such that $g * s = s'$.

(ii) Given s, s' and $v \in S$ such that $s' = g * s$, find $v' = g * v$.

To build a Diffie-Hellman protocol in this setting, first, publish a fixed element $s_0 \in S$. Alice and Bob can take random elements a and b from G for private keys. Then, their public keys are $a * s_0$ and $b * s_0$ and the shared secret is $b * (a * s_0) = a * (b * s_0)$. The security of private keys is based on the difficulty of the hard problem (i), while the security of the shared secret is protected by the hard problem (ii) itself.

Endomorphism ring

Definition 2.3.7. Let E be an elliptic curve over a field k. The set of all endomorphisms of the curve E, including the multiplication by 0 map, forms a ring, with addition and composition as operations, denoted by $\mathrm{End}(E)$.

The k-rational endomorphism ring is the subring $\mathrm{End}_k(E)$. When k is a finite field $\mathbb{F}_p$, the subring is denoted by $\mathrm{End}_p(E)$.

Remark. For any integer N, the multiplication by N map is an endomorphism, so $\mathbb{Z} \subset \mathrm{End}(E)$. The Frobenius map π, introduced in Definition 1.2.31, is an endomorphism, thus $\mathbb{Z}[\pi] \subset \mathrm{End}(E)$.

Definition 2.3.8. Let K be a finitely generated $\mathbb{Q}$-algebra. A subring $\mathcal{O}$ of K that is a finitely generated $\mathbb{Z}$-module[4] of maximal dimension is called an order.

Definition 2.3.9. A quaternion algebra is an algebra[5] of the form

$$K = \mathbb{Q} + \alpha\mathbb{Q} + \beta\mathbb{Q} + \alpha\beta\mathbb{Q},$$

where

$$\alpha^2\beta^2 \in \mathbb{Q}, \quad \alpha^2 < 0, \quad \beta^2 < 0, \quad \beta\alpha = -\alpha\beta.$$

An excellent reference for all things related to quaternions algebras is [68]. Quaternion algebras are an important part of the Deuring theorem (correspondence). This theorem provides a way to translate hard problems for elliptic curves to problems regarding maximal orders in quaternion algebras, which gives us another angle and more ”tools” to work on elliptic curve problems.

Theorem 2.3.10 (Deuring)**.** Let E be an elliptic curve defined over a field k of characteristic p. The ring $\mathrm{End}(E)$ is isomorphic to one of the following:

(i) $\mathbb{Z}$, only if $p = 0$.

(ii) A maximal order in the quaternion algebra ramified at p and 1, only if $p > 0$. In this case, we say that E is supersingular.

(iii) An order $\mathcal{O}$ in a quadratic imaginary field[6], for $p \geq 0$. In this case, we say that E has complex multiplication by $\mathcal{O}$.

Proof. See [64, Chapter III, Corollary 9.4.]. ■

The following theorem gives us an interesting property of the Frobenius endomorphism for an elliptic curve over a finite field.

Theorem 2.3.11. Let E be an elliptic curve defined over a finite field $\mathbb{F}_q$. The Frobenius endomorphism $\pi\colon E \to E$ satisfies a quadratic (characteristic) equation,

$$\pi^2 - t\pi + q = 0, \tag{2.1}$$

for some $|t| \leq 2\sqrt{q}$.

[4] A module can be viewed as a generalization of a vector space, where the field is replaced by a ring.

[5] Module together with a bilinear map.

[6] A number field of the form $\mathbb{Q}(\sqrt{-D})$ for some $D > 0$.

Proof. See [64, Chapter V, Theorem 2.3.1.]. ■

The coefficient t in (2.1) is called the trace of Frobenius. In a finite field with q elements, as we noted in Section 1.2.3, group $E(\mathbb{F}_q)$ is equal to $\ker(\pi - 1)$, from which $\#E(\mathbb{F}_q) = \#\ker(\pi - 1) = q - 1 + t$.

Let $\mathcal{O}$ be an order in a quadratic number field k. The norm of an $\mathcal{O}$-ideal $\mathfrak{a} \subseteq \mathcal{O}$ is defined as $N(\mathfrak{a}) = |\mathcal{O}/\mathfrak{a}|$. A fractional ideal of $\mathcal{O}$ is an $\mathcal{O}$-submodule of k of the form $\alpha\mathfrak{a}$, where $\alpha \in k^*$ and $\mathfrak{a}$ is an $\mathcal{O}$-ideal. The set of invertible fractional ideals $I(\mathcal{O})$ forms an abelian group under ideal multiplication. This group contains the principal fractional ideals $P(\mathcal{O})$ as a subgroup. The ideal-class group of $\mathcal{O}$ is the quotient

$$\mathrm{cl}(\mathcal{O}) = I(\mathcal{O})/P(\mathcal{O}).$$

Let $\pi \in \mathcal{O}$. The set of elliptic curves E defined over $\mathbb{F}_p$ with $\mathrm{End}_p(E) \cong \mathcal{O}$ such that π corresponds to the $\mathbb{F}_p$-Frobenius endomorphism of E is denoted with $\mathrm{Ell}_p(\mathcal{O}, \pi)$. Let $E \in \mathrm{Ell}_p(\mathcal{O}, \pi)$ and let $\mathfrak{a}$ be an integral ideal of $\mathcal{O}$. The $\mathfrak{a}$-torsion subgroup $E[\mathfrak{a}]$ of E is defined by

$$E[\mathfrak{a}] \coloneqq \{P \in E \colon [\psi]P = \mathcal{O}, \text{ for all } \psi \in \mathfrak{a}\}.$$

This is a finite subgroup, since $E[\mathfrak{a}] \subseteq E[N(\mathfrak{a})]$. There exist a unique elliptic curve $E/E[\mathfrak{a}]$ and an isogeny $\varphi_{\mathfrak{a}} \colon E \to E/E[\mathfrak{a}]$ with kernel $E[\mathfrak{a}]$. We denote the elliptic curve $E/E[\mathfrak{a}]$ by $\mathfrak{a} * E$. This correspondence induces an action, i.e. the ideal-class group of an imaginary quadratic order $\mathcal{O}$ acts freely via isogenies on the set of elliptic curves with $\mathbb{F}_p$-rational endomorphism ring isomorphic to $\mathcal{O}$. This is shown in the following theorem.

Theorem 2.3.12. Let $\mathcal{O}$ be an order in an imaginary quadratic field and $\pi \in \mathcal{O}$ such that $\mathrm{Ell}_p(\mathcal{O}, \pi)$ is nonempty. The ideal-class group $\mathrm{cl}(\mathcal{O})$ acts freely and transitively on the set $\mathrm{Ell}_p(\mathcal{O}, \pi)$ via the map

$$\begin{aligned} \mathrm{cl}(\mathcal{O}) \times \mathrm{Ell}_p(\mathcal{O}, \pi) &\to \mathrm{Ell}_p(\mathcal{O}, \pi) \\ ([\mathfrak{a}], E) &\mapsto \mathfrak{a} * E, \end{aligned}$$

in which $\mathfrak{a}$ is chosen as an integral representative.

Proof. See [18, Theorem 7.] and [55, Section 2.3.]. ■

A set that is acted upon freely and transitively by a group is called a principal homogeneous space for that group. For the Couveignes-Rostovtsev-Stolbunov key exchange, the space of public keys is equal to the set of $\mathbb{F}_q$-isomorphism classes of ordinary elliptic curves over $\mathbb{F}_q$ whose endomorphism ring is a given order $\mathscr{O}$ in an imaginary quadratic field and whose trace of Frobenius has a prescribed value.

CSIDH

Informally, CSIDH is the adaptation of the CRS scheme to supersingular elliptic curves defined over a field $\mathbb{F}_p$, where p is prime. The following is a description of a non-interactive[7] key exchange CSIDH:

- Public parameters of the system are a large prime

  $$p = 4 \cdot l_1 \cdots l_n - 1,$$

  where the l_i are small distinct odd primes, and the supersingular elliptic curve

  $$E_0 \colon y^2 = x^3 + x$$

  over $\mathbb{F}_p$ with endomorphism ring $\mathscr{O} = Z[\pi]$.

- **Key creation**:
  The private key is an integer n-tuple $(e_1, \ldots, e_n)$, where each integer is randomly sampled from a range $\{-m, \ldots, m\}$. These integers represent the ideal class

  $$[\mathfrak{a}] = [\mathfrak{l}_1^{e_1} \cdots \mathfrak{l}_n^{e_n}] \in \operatorname{cl}(\mathscr{O}),$$

  where $\mathfrak{l}_i = (l_1, \pi - 1)$.
  The public key is the coefficient $A \in \mathbb{F}_p$ of the elliptic curve

  $$[\mathfrak{a}]E_0 \colon y^2 = x^3 + Ax^2 + x$$

  obtained by applying the action of $[\mathfrak{a}]$ to the curve E_0.

- **Key exchange**:
  Let $(\mathfrak{a}, A)$ and $(\mathfrak{b}, B)$, $B \in \mathbb{F}_p \backslash \{\pm 2\}$ denote Alice and Bob's key pairs respectively.

[7] Two parties can exchange a symmetric key without any interaction.

Using Bob's public key B, Alice verifies that the elliptic curve $E_B\colon y^2 = x^3 + Bx^2 + x$ is indeed in $\mathrm{Ell}_p(\mathcal{O}, \pi)$. Alice then applies the action of her secret key $[\mathfrak{a}]$ to E_B to compute the curve

$$[\mathfrak{a}]E_B = [\mathfrak{a}][\mathfrak{b}]E_0.$$

Analogously, Bob proceeds to compute the curve $[\mathfrak{b}]E_A = [\mathfrak{b}][\mathfrak{a}]E_0$.

The shared secret is the coefficient S of the common secret curve $[\mathfrak{a}][\mathfrak{b}]E_0 = [\mathfrak{b}][\mathfrak{a}]E_0$ written in the form $y^2 = x^3 + Sx^2 + x$. This is the same for Alice and Bob since $\mathrm{cl}(\mathcal{O})$ is commutative.

Small primes l_i in the public parameter $p = 4 \cdot l_1 \cdots l_n - 1$, are chosen as Elkies primes, i.e. every ideal $l_i\mathcal{O}$ splits as $l_i\mathcal{O} = \mathfrak{l}_i\bar{\mathfrak{l}}_i$, where $\mathfrak{l}_i = (l_i, \pi - 1)$ and $\bar{\mathfrak{l}}_i = (l_i, \pi + 1)$ is a conjugate. More about design choices can be found in [18, Section 4].

Exponents e_i are, for efficiency, chosen from the small interval $\{-m, \ldots, m\}$, where m is selected such that $2m + 1 \geq \sqrt[n]{\#\mathrm{cl}(\mathcal{O})}$, see [18, Section 7.1.].

Remark. The security of the CSIDH scheme is an analogue to the security of the discrete-logarithm problem defined in Section 2.1.1. Let E, E' be two supersingular elliptic curves defined over $\mathbb{F}_p$ with the same $\mathbb{F}_p$-rational endomorphism ring $\mathcal{O}$. The DLP in this setting is to find an ideal $\mathfrak{a}$ of $\mathcal{O}$ such that $[\mathfrak{a}]E = E'$. If found, this ideal must be represented in such a way that the action of $[\mathfrak{a}]$ on a curve can be evaluated efficiently.

Remark. This thesis will not delve into details of the other well-known isogeny-based scheme SIDH. We are just going to briefly mention that since SIDH was first introduced, one of the most questionable parts of the scheme has been that publishing images of known points under secret isogenies could be a way to break the scheme. In 2022, this became evident in several attacks [15, 47, 60]. Both CRS and CSIDH are not publishing such additional points or images and are not affected by these attacks.

3. RADICAL ISOGENIES

This chapter provides the background on radical isogenies; formulas designed for calculating isogenies between elliptic curves. We will mostly follow [17] and [54].

3.1. DEFINITION OF RADICAL ISOGENIES

Let k be a field, and $N \geq 4$ an integer such that $\mathrm{char}(k) \nmid N$. Consider an elliptic curve E over k and a point $P \in E(k)$ of order N. Using Lemma 1.2.8, the curve-point pair (E,P) is isomorphic to a unique pair of a curve

$$y^2 + (1-c)xy - by = x^3 - bx^2,$$

where $b,c \in k$, and a point $(0,0)$ of order N.

Remark. In [17] elliptic curve E is defined over field

$$\mathbb{Q}_N(b,c) := \mathrm{Frac}\frac{\mathbb{Q}[b,c]}{(F_N(b,c))},$$

where "Frac" is a field of fractions[1], and F_N is a polynomial defined by (1.4). The field $\mathbb{Q}_N(b,c)$ is the function field of $X_1(N)$ over $\mathbb{Q}$.

According to Theorem 1.2.18, there exists an isogeny

$$\varphi\colon E \to E/\langle P\rangle$$

with $\langle P\rangle$, a cyclic subgroup generated by the point P, as a kernel. We denote the curve $E/\langle P\rangle$ over k by E'. We are interested in the points $P' \in E'$ for which the composition

$$E \xrightarrow{\varphi} E' \to E'/\langle P'\rangle$$

[1] A field of fractions of an integral domain R (nonzero commutative ring in which the product of any two nonzero elements is nonzero) is the smallest field containing R.

is cyclic N^2-isogeny, satisfying the condition

$$\widehat{\varphi}(P') = [\lambda]P, \quad \text{for some} \quad \lambda \in (\mathbb{Z}/N\mathbb{Z})^{\times},$$

where $\widehat{\varphi}$ is a dual isogeny of φ. This condition is valid for $N\phi(N)$ points[2]. A nonunique point P' corresponding to $\lambda = 1$ is called P-distinguished. Radical isogenies are formulas for the coordinates of the point P'.

Let $f_{N,P}$ be the normalized Miller function and define

$$\rho := f_{N,P}(-P).$$

From Section 1.2.2, when the Miller function is normalized, Tate pairing $t_N(P,-P)$ can be calculated as $f_{N,P}(-P)$. The following theorem describes the field of definition of the point P'.

Theorem 3.1.1. Let E be an elliptic curve defined over the field $\mathbb{Q}_N(b,c)$ and P a point of order N on that curve. Let $\varphi\colon E \to E' := E/\langle P\rangle$ be an isogeny with a kernel equal to $\langle P\rangle$, and $P' \in E'$ a point of order N satisfying $\widehat{\varphi}(P') = [\lambda]P$. The field extension $\mathbb{Q}_N(b,c) \subset \mathbb{Q}_N(b,c)(P')$, obtained by adjoining the coordinates of P', is a simple radical extension of degree N. More precisely, for an appropriately chosen $\sqrt[N]{\rho}$ of $\rho = f_{N,P}(-P)$, where $f_{N,P}$ is a normalized Miller function,

$$\mathbb{Q}_N(b,c)(P') = \mathbb{Q}_N(b,c)(\sqrt[N]{\rho}).$$

Proof. See [17, Theorem 5.]. ■

According to Theorem 3.1.1, the coordinates of the point P' can be calculated using a rational expression that depends on b,c and $\sqrt[N]{\rho}$, where ρ is a representative of Tate pairing $t_N(P,-P)$. Hence, the point P' is defined over $k(b,c,\sqrt[N]{\rho})$. Formulas for all other points that are not P-distinguished can be calculated using multiplication by λ map and (1.3). Furthermore, a P-distinguished point P' is nonunique, but all other P-distinguished points are found by varying the choice of $\sqrt[N]{\rho}$, i.e. by scaling it with ζ_N^i, $i = 0,\ldots,N-1$, where ζ_N is a primitive N-root of unity.

As P' is of order N on the curve E', a Tate normal form for this pair can be defined by unique coefficients b' and c'. The iterative process of radical isogeny formulas can be

[2] $\phi(N)$ is the Euler function.

repeated on the pair (E', P'). Moreover, the formulas for b' and c' are expressed directly as elements of the field extension $k(b, c, \sqrt[N]{\rho})$, which is a simple radical extension of $k(b, c)$. Explicit radical isogeny formulas when $N = 5$, are given in the following example.

Example 3.1.2 ([17, Section 4.]). For $N \geq 4$, the point $P = (0, 0)$ on E is of order N if and only if $F_N = 0$. Let $N = 5$. From (1.4), polynomial $F_5(b, c)$ is equal to $c - b$, and in that case, elliptic curve E has a modified Tate normal form

$$y^2 + (1-b)xy - by = x^3 - bx^2.$$

Using Vélu's formulas, the curve E' is equal to

$$y^2 + (1-b)xy - by = x^3 - bx^2 - 5b(b^2 + 2b - 1)x - b(b^4 + 10b^3 - 5b^2 + 15b - 1).$$

The x-coordinates of the point P' are roots of 5-division polynomial ψ_5 for the curve E'. This polynomial splits into factors as

$$\begin{aligned}
\psi_{E',5}(x) = 5 \cdot & (x^2 + (b^2 - b + 1)x + (b^4 + 3b^3 - 26b^2 - 8b + 1)/5) \\
& \cdot (x^5 + 10bx^4 - 5b(b^2 + b - 11)x^3 - 5b(17b^3 + 24b^2 + 46b - 7)x^2 \\
& \quad - 5b(b^5 + 62b^4 + 154b^3 - 65b^2 + 19b - 2)x \\
& \quad - b(b^7 - 19b^6 + 777b^5 - 757b^4 + 755b^3 + 2b^2 + 17b - 1)) \\
& \cdot (x^5 - 15bx^4 - 5b(11b^2 - 9b - 1)x^3 - 5b^2(7b^3 + 13b^2 - 13b + 20)x^2 \\
& \quad - 5b^2(2b^5 + 5b^4 + 6b^3 + 196b^2 - 99b + 1)x \\
& \quad - b^2(b^7 + 7b^6 - 62b^5 + 605b^4 - 127b^3 + 1177b^2 + 14b + 1)).
\end{aligned}$$

The quadratic polynomial factor of $\psi_{E',5}$ is the kernel polynomial of the dual isogeny. The roots of the first quintic polynomial factor are the x-coordinates we are interested in, and the roots of the second quintic polynomial factor are the x-coordinates of points P'' satisfying $\widehat{\varphi}(P'') = 2P$. If we let $\rho = f_{5,P}(-P) = b$ and denote $\alpha = \sqrt[5]{\rho}$, the coordinates of point P' are

$$\begin{aligned}
x'_0 &= 5\alpha^4 + (b-3)\alpha^3 + (b+2)\alpha^2 + (2b-1)\alpha - 2b, \\
y'_0 &= 5\alpha^4 + (b-3)\alpha^3 + (b^2 - 10b + 1)\alpha^2 + (13b - b^2)\alpha - b^2 - 11b.
\end{aligned}$$

After translating the point P' to $(0, 0)$, the isomorphic curve in Tate normal form is

$$E' : y^2 + (1 - b')xy - b'y = x^3 - b'x^2,$$

where

$$b' = \alpha \frac{\alpha^4 + 3\alpha^3 + 4\alpha^2 + 2\alpha + 1}{\alpha^4 - 2\alpha^3 + 4\alpha^2 - 3\alpha + 1}$$

and the process can be repeated.

The standard method of calculating isogenies (Vélu's formulas) would require a point of particular order for each isogeny in the chain. With radical isogeny formulas, such a point is only required for the initial step, i.e. the step that uses Vélu's formulas. Subsequent steps can be calculated without any knowledge of torsion points. The list of formulas for radicand ρ for $N \leq 13$ can be found in [17, Section 5.]. The link to a repository containing formulas for prime powers $16 < N \leq 37$ can be found in [16, Section 4.3.].

In Chapter 2 we mentioned the protocol CSIDH, one of the isogeny-based key exchange protocols. Computing a composition of a large number of isogenies is an important step in CSIDH. Radical isogenies have a direct application in this protocol and using them can lead to a speed-up of about 35% for CSIDH-512. See [16, Section 7.2.] for more details.

3.2. RADICAL ISOGENIES ON MONTGOMERY CURVES

Radical isogenies formulas are not restricted to curves in Tate normal form, and they can be defined for other curve models. This section is dedicated to radical isogenies on Montgomery curves and it mostly follows [54]. First, some background on Montgomery curves.

3.2.1. Montgomery curves

This section mostly follows [38, Part II, Section 9.12.] and [24]. Montgomery curves were first introduced in [52] by Peter Montgomery (hence the name) as a tool to accelerate some previously known factorization methods. We will start with the definition of the Montgomery curve.

Definition 3.2.1. Let k be a field such that $\mathrm{char}(k) \nmid 2$. A Montgomery curve over a field k is an elliptic curve of the form

$$E: By^2 = x^3 + Ax^2 + x, \tag{3.1}$$

where $A, B \in k$ and $B(A^2-4) \neq 0$.

Remark. The condition $B(A^2-4) \neq 0$ ensures that the curve is nonsingular.

There is a group law on the points of E. Let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ be points on the elliptic curve

$$E: By^2 = x^3 + Ax^2 + x,$$

such that $x_1 \neq x_2$ and $x_1 x_2 \neq 0$. Then $P_1 + P_2 = (x_3, y_3)$, where

$$x_3 = B\frac{(x_2y_1 - x_1y_2)^2}{x_1x_2(x_2 - x_1)^2}.$$

For the case $P_2 = P_1$ we have $[2](x_1, y_1) = (x_3, y_3)$, where

$$x_3 = \frac{(x_1^2 - 1)^2}{4x_1(x_1^2 + Ax_1 + 1)}.$$

There is a connection between curves given in the Weierestrass form and curves in the Montgomery form.

Lemma 3.2.2. Let k be a field such that $\text{char}(k) \neq 2$. Let

$$E: y^2 = x^3 + a_2x^2 + a_4x + a_6$$

be an elliptic curve over k in the Weierstrass form. There is an isomorphism over k from E to a Montgomery model if and only if

$$F(x) = x^3 + a_2x^2 + a_4x + a_6$$

has a root $x_P \in k$ such that $(3x_P^2 + 2a_2x_P + a_4)$ is a square in k. This isomorphism maps $\mathcal{O}_E$ to the point at infinity on the Montgomery model and it is a group homomorphism.

Proof. See [38, Part II, Lemma 9.12.4.]. ■

In Section 2.1.1, we discussed the elliptic curve Diffie-Helman key exchange and mentioned that one of the earliest proposals for using elliptic curves in cryptography was made by Miller in [51]. In the same paper, Miller presented a method for performing the Diffie-Helman key exchange using only the x-coordinate of a public point P. If we denote by $x(P)$ the x-coordinate of a point P, then it can be shown that Alice can calculate the shared secret $x([n_A]([n_B]P)) = x([n_B]([n_A]P))$ from $x([n_B]P)$ and Bob can do the same from $x([n_A]P)$. Working only with the x-coordinate can offer time and space optimization in the implementation, providing that we use a class of curves with efficient algorithms for computing point multiplication $x(P) \mapsto x([k]P)$, for some integer k. This is where Montgomery curves are considered to be the optimal choice, particularly due to the effectiveness of the Montgomery ladder algorithm. The Montgomery ladder offers improved security and efficiency compared to the other standard algorithms with the same purpose (the double-and-add algorithm). For a detailed description of the Montgomery ladder, we refer to [24, Chapter 4.].

3.2.2. Definition of radical isogenies on Montgomery curves

In this section, we will follow [54] to explain radical isogenies on Montgomery curves of degrees 3 and 4. Montgomery curves are used in an effort to reduce the computation cost of radical isogenies. Throughout this section, we let E to be a Montgomery curve over a

field k, where $\mathrm{char}(k) \neq 2$, and the coefficient B is equal to 1, i.e. an elliptic curve of the form

$$E\colon y^2 = x^3 + Ax^2 + x,$$

where $A \in k$, and $A^2 \neq 4$. Commonly used notation for a point $P = (x,y)$ on a Montgomery curve E is $(x,-)$, where the y-coordinate is in most cases implied or reconstructed. The j-invariant of a curve E is equal to

$$j(E) = \frac{256(A^2-3)^3}{A^2-4}.$$

The formula for the j-invariant implies that there are six isomorphic Montgomery curves over $\overline{k}$ (remember that the curves are isomorphic over $\overline{k}$ if they have the same j-invariant). Even more, the coefficient A determines a class of the enhanced elliptic curve $(E,(0,0))$ in the set $S_0(4)$, as given in the following proposition.

Proposition 3.2.3. Let E and E' be two Montgomery curves over k with A and A' as their Montgomery coefficients, respectively. Let $C_E^{(4)}$ denote the cyclic subgroup of E generated by $(1,-) \in E(\overline{k})$. Then $(E,C_E^{(4)}) \sim (E',C_{E'}^{(4)})$ if and only if $A = A'$. Furthermore, $(E,\langle(0,0)\rangle) \sim (E',\langle(0,0)\rangle)$ if and only if $A^2 = A'^2$.

Proof. See [54, Proposition 1.]. ■

Let (E,C) be an enhanced elliptic curve for $\Gamma_0(4)$ over k. One can show that there exists a Montgomery curve E' and an isomorphism $E \to E'$ that takes C to $C_{E'}^{(4)}$. Therefore, we can define a bijection from the set $S_0(4)$ to $\overline{k}\backslash\{\pm 2\}$ by sending a class $[E,C]$ to the Montgomery coefficient of a Montgomery curve in the class $[E,C]$.

Applying radical isogeny formulas on elements of set $S_1(N)$, i.e. on an enhanced elliptic curve (E,P), results in a curve-point pair that is also an element of $S_1(N)$. When $N = 3$ or 4, the modular curves $X_0(N)$ and $X_1(N)$ are isomorphic, i.e. there is a bijection between sets $S_0(N)$ and $S_1(N)$, so the existence of radical isogeny formulas on $S_1(3)$ and $S_1(4)$ implies a radical isogeny formula on $S_0(3)$ and $S_0(4)$, respectively. This means that there is a formula between the Montgomery coefficients of curves. For $N = 4$ we have the following theorem (for $N = 3$, see [54, Section 3.1.]).

Theorem 3.2.4. Let E be a Montgomery curve with coefficient $A \in k$, E' a Montgomery curve, map $\varphi\colon E \to E'$ an isogeny with kernel $C_E^{(4)}$, and ψ an isogeny from E' with kernel

$\langle(0,0)\rangle$. If the kernel of the composition $\psi\circ\varphi$ is cyclic, the Montgomery coefficient A' of E' is

$$\frac{(\beta+2)^4}{4\beta(\beta^2+4)}-2,$$

where β is a fourth root of $4(A+2)$.

Proof. See [54, Theorem 8.]. ■

The methods used in [54] for cases $N=3$ or 4 cannot be directly applied to the case $N\geq 5$, partly because sets $S_0(N)$ and $S_1(N)$ are not bijective. Moreover, developing radical isogeny formulas on $S_0(N)$ when $N\geq 5$ might not be possible, as illustrated by the following example.

Example 3.2.5 ([54, Section 4.]). Let $N=5$. Let k be a field with $\mathrm{char}(k)\nmid N$, and let E and E' be two elliptic curves over the field k given in Tate normal form:

$$E\colon y^2+(1-b)xy-by=x^3-bx,$$
$$E'\colon y^2+(1-b')xy-b'y=x^3-b'x.$$

The points $(0,0)$ are of order 5 on these curves. The cyclic subgroup of E generated by point $(0,0)$ is

$$\{\mathcal{O}_E,(0,0),(b,b^2),(b,0),(0,b)\}.$$

The pairs $(E,(0,0))$ and $(E',(0,0))$ are equivalent if and only if $b=b'$, while the pairs $(E,\langle(0,0)\rangle)$ and $(E',\langle(0,0)\rangle)$ are equivalent if and only if $b=b'$ or $b=-\frac{1}{b'}$. From this we have $\frac{b^2-1}{b}=\frac{b'^2-1}{b'}$, thus $\frac{b^2-1}{b}$ is a parametrization of $S_0(5)$. If the curves E and E' are isogenous, from the radical isogeny formula we know that b' is a rational expression in the fifth root of b, i.e. $\mathbb{Q}(b')=\mathbb{Q}(\sqrt[5]{b})$. Let $\beta=\frac{b^2-1}{b}$ and $\beta'=\frac{b'^2-1}{b'}$. Field extension $\mathbb{Q}(b)/\mathbb{Q}(\beta)$ is of degree 2. If we adjoin to the field extension $\mathbb{Q}(b')/\mathbb{Q}(\beta)$ a primitive fifth root of unity $\zeta_5\in\mathbb{C}$, we obtain a Galois extension $\mathbb{Q}(\zeta_5)(b')/\mathbb{Q}(\zeta_5)(\beta)$ of degree 10. The Galois group of this extension $\mathrm{Gal}(\mathbb{Q}(\zeta_5)(b')/\mathbb{Q}(\zeta_5)(\beta))$ is generated by automorphisms $\sigma\colon b'\mapsto-\frac{1}{b'}$ and $\tau\colon b'\mapsto\zeta_5 b'$. The fixed field of σ is $\mathbb{Q}(\zeta_5)(\beta')$, and the fixed field of τ is $\mathbb{Q}(\zeta_5)(b)$. Because $\tau^{-1}\sigma\tau\neq\sigma$, the group $\langle\sigma\rangle$ is not a normal subgroup of Galois group $\mathrm{Gal}(\mathbb{Q}(\zeta_5)(b')/\mathbb{Q}(\zeta_5)(\beta))$, thus the extension $\mathbb{Q}(\zeta_5)(\beta')/\mathbb{Q}(\zeta_5)(\beta)$ cannot be a Galois extension.

If the parameter β' from Example 3.2.5 could be expressed as a rational expression depending on the parameter β, we would have a way to calculate b' (quadratic equation) that is simpler and more direct than radical isogeny formulas. However, this is not possible, because the field extension $\mathbb{Q}(\zeta_5)(\beta')/\mathbb{Q}(\zeta_5)(\beta)$ is not a Galois extension. Nevertheless, it may be possible to find a different β', i.e. a different parametrization of $S_0(5)$ which will make the field extension $\mathbb{Q}(\zeta_5)(\beta')/\mathbb{Q}(\zeta_5)(\beta)$ Galois. We discuss this further in Chapter 5.

4. RADICAL ISOGENIES IN THE LANGUAGE OF MODULAR CURVES - THE CASE $X_1(N)$

The goal of this chapter is to show that radical isogenies can be generalized using modular curves. To this end, we will be using enhanced elliptic curves for different congruence subgroups and maps between them. The notation used in this (and the following) chapter is the same as the one introduced in Chapter 3. Additionally, we will assume that ζ_N is an element of the field k. This assumption will be explained in Section 4.2 at the end of this chapter. The elliptic curve E will be the starting elliptic curve over a field k, $N \geq 4$ such that $\mathrm{char}(k) \nmid N$, $P \in E(k)$ a point of order N, E' a curve over k isomorphic to $E/\langle P \rangle$, $\varphi\colon E \to E'$ an isogeny with kernel equal to $\langle P \rangle$ and P' a point of order N on E' such that $\widehat{\varphi}(P') = P$.

4.1. GENERALIZATION OF RADICAL ISOGENIES

For any elliptic curve $\widetilde{E}$ and point $\widetilde{P}$ of order $N \geq 4$, let its unique Tate normal form be defined with parameters $\widetilde{b}$ and $\widetilde{c}$. Let $\mathbf{b}$ denote a mapping

$$\mathbf{b}\colon (\widetilde{E}, \widetilde{P}) \mapsto \widetilde{b},$$

i.e. $\mathbf{b}$ is a function on the set of enhanced elliptic curves for $\Gamma_1(N)$, such that for a curve $(\widetilde{E}, \widetilde{P})$ it returns the parameter $\widetilde{b}$ from the corresponding Tate normal form. This is a well-defined function because the Tate normal form is unique. Analogously, for a parameter $\widetilde{c}$, the function $\mathbf{c}\colon (\widetilde{E}, \widetilde{P}) \mapsto \widetilde{c}$ is a well-defined function. The definition of modular functions

on enhanced elliptic curves implies that **b** and **c** are elements of $k(X_1(N))$. For curves E and E' we have four such mappings

$$\begin{aligned}(E,P) &\overset{\mathbf{b}}{\mapsto} b \quad \text{and} \quad (E,P) \overset{\mathbf{c}}{\mapsto} c \quad \text{for } E,\\ (E',P') &\overset{\mathbf{b}}{\mapsto} b' \quad \text{and} \quad (E',P') \overset{\mathbf{c}}{\mapsto} c' \quad \text{for } E'.\end{aligned}$$

We would like to connect parameters b,c with b',c' using modular curves and maps between them. The following sequence of maps will be considered:

$$\begin{aligned}(E,P) &\to (E',P') \overset{\mathbf{b}}{\mapsto} b',\\ (E,P) &\to (E',P') \overset{\mathbf{c}}{\mapsto} c'.\end{aligned} \tag{4.1}$$

Since the point P' is not unique, the map $(E,P) \to (E',P')$ is not uniquely defined, and therefore no obvious connection on $X_1(N)$ exists. For a point P of order N, let R be a point on the curve E of order N^2 such that $[N]R = P$. This point R is not unique. The pair (E,R) is an enhanced elliptic curve for $\Gamma_1(N^2)$. Let P' be an image of a point R under the isogeny φ, i.e.

$$P' := \varphi(R) = R + \langle P \rangle.$$

This is a point of order N on the curve E'.

Remark. The point P' is of order N on curve E' since

$$[N]P' = [N]R + \langle P \rangle = P + \langle P \rangle = \mathcal{O}_{E'},$$

where the last equality is valid because we have

$$\overline{P} + \langle P \rangle = \mathcal{O}_{E'} \Leftrightarrow \overline{P} \in \langle P \rangle, \quad \forall \overline{P} \in E'(\bar{k}).$$

Additionally,

$$\widehat{\varphi}(P') = \widehat{\varphi}(\varphi(R)) = [\deg(\varphi)]R = [N]R = P,$$

so point P' is also P-distinguished. We can modify the sequence of maps in (4.1) and continue to work with the parameter b and associated functions, as the approach for c is the same. Beginning with the enhanced elliptic curve (E,R), we have the following maps:

$$(E,R) \to (E,[N]R) = (E,P) \overset{\mathbf{b}}{\mapsto} b, \tag{4.2}$$

$$(E,R) \to (E/\langle [N]R \rangle, R + \langle [N]R \rangle) = (E/\langle P \rangle, R + \langle P \rangle) = (E',P') \overset{\mathbf{b}}{\mapsto} b'. \tag{4.3}$$

Using the mappings described in (4.3), we can, similar to $\mathbf{b}$, define a function

$$\mathbf{b}' \colon (E,R) \mapsto b',$$

which is a function on the set of enhanced elliptic curves for $\Gamma_1(N^2)$. Maps and functions are visualized in Figure 4.1.

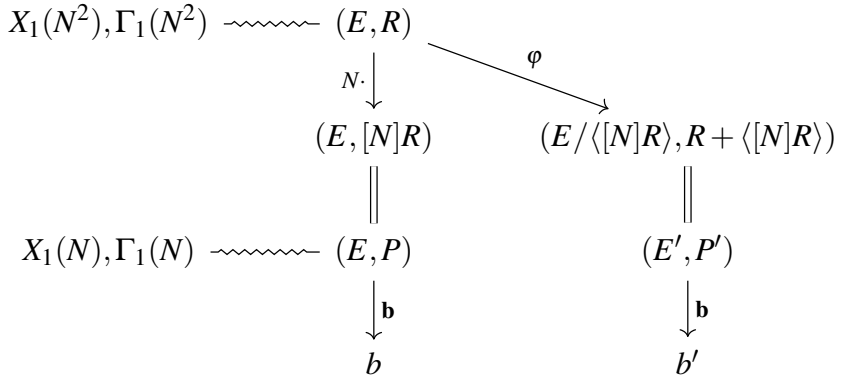

Figure 4.1: Maps between enhanced elliptic curves

The connection between parameters b and b' can now be extended to an enhanced elliptic curve (E,R), i.e. to functions on $X_1(N^2)$. For every N, let $\pi_{1,N}^*$ and $\pi_{2,N}^*$ define a pair of pullback operators:

$$\begin{aligned}
&\pi_{1,N}^* \colon k(X_1(N)) \to k(X_1(N^2)), \text{ of the map } \pi_{1,N}((E,R)) = (E,[N]R) \text{ and} \\
&\pi_{2,N}^* \colon k(X_1(N)) \to k(X_1(N^2)), \text{ of the map } \pi_{2,N}((E,R)) = (E/\langle[N]R\rangle, R+\langle[N]R\rangle).
\end{aligned} \tag{4.4}$$

From

$$(\pi_{1,N}^*\mathbf{b})(E,R) = \mathbf{b}(\pi_{1,N}((E,R))) = \mathbf{b}(E,[N]R) = \mathbf{b}(E,P)$$

and

$$(\pi_{2,N}^*\mathbf{b})(E,R) = \mathbf{b}(\pi_{2,N}((E,R))) = \mathbf{b}(E/\langle[N]R\rangle, R+\langle[N]R\rangle) = \mathbf{b}(E',P') = \mathbf{b}'(E,R),$$

we can identify $\mathbf{b}$ and $\mathbf{b}'$ with their respective pullbacks by $\pi_{1,N}$ and $\pi_{2,N}$ and define

$$b \coloneqq \pi_{1,N}^*\mathbf{b} \quad \text{and} \quad b' \coloneqq \pi_{2,N}^*\mathbf{b}$$

as functions on $X_1(N^2)$.

The function b' is an element of the field $\pi_{2,N}^*(k(X_1(N)))$. In order to generalize radical isogenies using modular curves, as b is an element of the field $\pi_{1,N}^*(k(X_1(N)))$, the function b' needs to be an element of some extension field of $\pi_{1,N}^*(k(X_1(N)))$. For this to be true, we need to find a function g in $k(X_1(N^2))$ such that

$$\pi_{1,N}^*(k(X_1(N)))(g) = \pi_{2,N}^*(k(X_1(N))). \tag{4.5}$$

Let P be a point of order N as before, and let $f_{N,P}$ be a normalized Miller function. With the value of $f_{N,P}$ at point $-P$, we can define a modular function f on the set of enhanced elliptic curves for $\Gamma_1(N)$ as:

$$f\colon (E,P) \mapsto f_{N,P}(-P) \in k(X_1(N)).$$

For the function $f_{N,P}$ and the point P, from equation (1.5), there exists a function $g_{N,P} \in \overline{k}(E)$ such that

$$f_{N,P} \circ [N] = g_{N,P}^N.$$

Furthermore, from the discussion after the equation (1.5), we can choose functions $f_{N,P}$ and $g_{N,P}$ to be in $k(E)$. Using so chosen functions and the equality between them, for an enhanced elliptic curve (E,R), where, as before $P = [N]R$, we have a function on $X_1(N^2)$ given by

$$\begin{aligned}(E,R) \mapsto f_{N,[N]R}(-[N]R) &= f_{N,[N]R}([N](-R)) \\ &= g_{N,[N]R}(-R)^N = g_{N,P}(-R)^N.\end{aligned}$$

The function g defined as

$$g \coloneqq (E,R) \mapsto g_{N,P}(-R)$$

is an element of the field $k(X_1(N^2))$ and it is satisfying the property

$$g^N = f,$$

which means that the N-th root of f is a function on $X_1(N^2)$. Both functions b, b', as well as function g are elements of $k(X_1(N^2))$. However, due to the large size of this field, it is currently impossible to prove the equality (4.5). Thus, it is necessary to identify a smaller quotient of $X_1(N^2)$ where b, b', and g are well-defined.

4.1.1. Finding the quotient

To gain a better understanding of the function b', we will investigate the preimages of (E,P) under the pullback operator $\pi_{2,N}$. Specifically, we will investigate pairs (E,R) and (E,R') that are mapped by $\pi_{2,N}$ to the same point $(E/\langle [N]R\rangle, R+\langle [N]R\rangle)$. For the equality

$$(E/\langle [N]R'\rangle, R'+\langle [N]R'\rangle) = (E/\langle [N]R\rangle, R+\langle [N]R\rangle)$$

to hold, we require $\langle [N]R'\rangle = \langle [N]R\rangle$ and $R'+\langle [N]R'\rangle = R+\langle [N]R\rangle$. Combining these conditions, we get $R'+\langle [N]R\rangle = R+\langle [N]R\rangle$, which implies that there exists some $l \in \mathbb{Z}/N\mathbb{Z}$ such that

$$R' = R+[l]\cdot([N]R) \text{ and } [N]R' = [N](R+[l]P).$$

Therefore, we have

$$\langle [N](R+[l]P)\rangle = \langle [N]R\rangle.$$

Since the point R has order N^2, the points $(E,R),(E,R+[1\cdot N]R),\ldots,(E,R+[(N-1)\cdot N]R)$ are all mapped to the same final point. From the definition of b', it is apparent that it is a function on $X_1(N^2)$ that maps points of this form to the same final point.

Let m be an integer that is relatively prime with N. Let t_m be an operator on $S_1(N^2)$ defined as $t_m\colon (E,\overline{P}) \mapsto (E,[m]\overline{P})$. When $m = N+1$, define $t := t_{N+1}$. On an enhanced elliptic curve $(E,R) \in S_1(N^2)$, this operator acts as:

$$(E,R) \overset{t}{\mapsto} (E,[N+1]R) \overset{t}{\mapsto} (E,[(N+1)^2]R) \overset{t}{\mapsto} \cdots \overset{t}{\mapsto} (E,[(N+1)^{N-1}]R).$$

The order of the operator t is equal to N since we have

$$t^N(E,R) = (E,[(N+1)^N]R) = (E,R).$$

Remark. In [31, Chapter 5, Section 5.2.], the moduli space diamond operator $\langle d\rangle$ is defined by

$$\langle d\rangle\colon S_1(N) \to S_1(N), \quad (E,P) \mapsto (E,[d]P), \quad (d,N) = 1.$$

The operator t is the diamond operator

$$\langle N+1\rangle\colon S_1(N^2) \to S_1(N^2), \quad (E,R) \mapsto (E,[N+1]R).$$

Composing t with $\pi_{1,N}$ on the enhanced elliptic curve (E,R), we have:

$$\begin{aligned}\pi_{1,N}(t(E,R)) &= \pi_{1,N}((E,[N+1]R))\\ &= (E,[N(N+1)]R) \text{ (since order of } R \text{ is } N^2)\\ &= (E,[N]R)\\ &= \pi_{1,N}(E,R),\end{aligned}$$

and for $\pi_{2,N}$:

$$\begin{aligned}\pi_{2,N}(t(E,R)) &= \pi_{2,N}((E,[N+1]R))\\ &= (E/\langle[N(N+1)]R\rangle,[N+1]R+\langle[N(N+1)]R\rangle)\\ &= (E/\langle[N]R\rangle,[N]R+R+\langle[N]R\rangle) \text{ (since order of } R \text{ is } N^2)\\ &= (E/\langle[N]R\rangle,R+\langle[N]R\rangle) \text{ (since } [N]R\in\langle[N]R\rangle)\\ &= \pi_{2,N}(E,R),\end{aligned}$$

thus, every pullback by $\pi_{1,N}$ or by $\pi_{2,N}$ will be invariant under the operator t.

The modular function $(E,R)\overset{g}{\mapsto} g_{N,P}(-R)$, with property $g^N=f$, is also invariant under t. From Section 1.2, we know that the function $g_{N,[N]R}$ can be used to define the Weil pairing. To see that the function $g\colon (E,R)\mapsto g_{N,[N]R}(-R)$ is invariant under t, let $(E,R)\in S_1(N^2)$. Then

$$\begin{aligned}t(E,R)=(E,[N+1]R) &\overset{g}{\mapsto} g_{N,[N(N+1)]R}(-[N+1]R)\\ &= g_{N,[N]R}(-[N]R-R)\\ &= g_{N,[N]R}(-R-P).\end{aligned}$$

Using the properties of the Weil pairing from Proposition 1.2.27, we have the following sequence of equalities:

$$\begin{aligned}g_{N,[N]R}(-R-P) &= g_{N,[N]R}(-R)e_N(-P,P) \text{ (definition of the Weil pairing with } S=-P)\\ &= g_{N,[N]R}(-R)e_N(-P,P)e_N(P,P)e_N(P,P) \text{ (alternating property)}\\ &= g_{N,[N]R}(-R)e_N(-P+P+P,P) \text{ (bilinearity)}\\ &= g_{N,[N]R}(-R)e_N(P,P)\\ &= g_{N,P}(-R).\end{aligned}$$

Let $\langle t \rangle$ denote the group of automorphisms of $X_1(N^2)$ generated by t. A function on $X_1(N^2)$ that is invariant under the operator t can be viewed as a function on the quotient $X_1(N^2)/\langle t \rangle$. As discussed above, b' is an example of such a function. The quotient $X_1(N^2)/\langle t \rangle$, i.e. the quotient of a modular curve by the operator t, is again a modular curve. To see this, following [30] and [42] we can assume, for a field k defined at the beginning of this chapter, that $k = \mathbb{C}$. Then, we have the following proposition, which explicitly calculates the congruence subgroup defining this quotient, i.e. the corresponding modular curve.

Proposition 4.1.1. Let t be an operator defined on the set of enhanced elliptic curves for $\Gamma_1(N^2)$ with $t(E,R) = (E,[N+1]R)$. Let $\langle t \rangle$ denote the subgroup of automorphisms of $X_1(N^2)$ generated by t. The quotient of the extended upper half-plane $\mathscr{H}^* = \mathscr{H} \cup \mathbb{Q} \cup \{\infty\}$ and the congruence subgroup

$$\widetilde{\Gamma}(N) \coloneqq \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \colon c \equiv 0 \pmod{N^2},\ a,d \equiv 1 \pmod{N} \right\},$$

i.e. $\widetilde{\Gamma}(N)/\mathscr{H}^*$, is a modular curve whose function field consists of all the functions on $X_1(N^2)$ invariant under t.

Proof. As shown in [31, Chapter 1, Section 1.5.], the sets of equivalence classes of enhanced elliptic curves can be used to describe quotients of the upper half-plane by congruence subgroups. In other words, for a function f on $X_1(N^2)/\langle t \rangle$, there is a corresponding meromorphic function $\mathbf{f}$ on the upper half-plane that is invariant under the action of $\Gamma_1(N^2)$ and a matrix $\mathbf{t} \in \mathrm{SL}_2(\mathbb{Z})$ corresponding to the operator t. To see this, note that Theorem 1.3.9 shows that $S_1(N^2)$ is a moduli space of isomorphism classes of complex elliptic curves and N^2-torsion data, i.e.

$$S_1(N^2) = \{[E_\tau, \frac{1}{N^2} + \Lambda_\tau]\},$$

where τ, Λ_τ and E_τ are defined as in Section 1.3. Describing what the operator t does in the sense of congruence subgroup implies working with the pair (E,R) after applying the operator t, i.e. with

$$t(E_\tau, \frac{1}{N^2} + \Lambda_\tau) = (E_\tau, \frac{N+1}{N^2} + \Lambda_\tau).$$

We need to find $\tau' \in \mathscr{H}$, such that $(E_\tau, \frac{N+1}{N^2} + \Lambda_\tau)$ is isomorphic to $(E_{\tau'}, \frac{1}{N^2} + \Lambda_{\tau'})$. In other words, we need to find an $\lambda \in \mathbb{C}$ such that the lattices Λ_τ and $\Lambda_{\tau'}$ are homothetic,

i.e. $\lambda\Lambda_{\tau'} = \Lambda_\tau$. Let

$$\tau' = \frac{(1-N)\tau - 1}{N^2\tau + 1 + N} \quad \text{and} \quad \Lambda_{\tau'} = \langle 1, \tau' \rangle.$$

Let $\lambda = N^2\tau + N + 1$. Elements 1 and τ are linear combinations (over $\mathbb{Z}$) of vectors $\lambda \cdot 1$ and $\lambda \cdot \tau'$, which is obvious for 1, and for τ we have:

$$(1+N) \cdot \lambda \cdot \tau' + \lambda \cdot 1 = \tau.$$

This implies $\Lambda_\tau \subset \lambda\Lambda_{\tau'}$. As τ' is, by definition, a linear combination (over $\mathbb{Z}$) of $\frac{\tau}{\lambda}$ and $\frac{1}{\lambda}$, we have $\Lambda_{\tau'} \subset \lambda\Lambda_\tau$. Therefore, Λ_τ and $\Lambda_{\tau'}$ are indeed homothetic.

Moreover, for the matrix

$$\mathbf{t} = \begin{pmatrix} 1-N & -1 \\ N^2 & 1+N \end{pmatrix} \in \Gamma_1(N) \setminus \Gamma_1(N^2),$$

using the usual fractional linear transformation on $\mathscr{H}$, we have $\mathbf{t}(\tau) = \tau'$. The desired congruence subgroup $\widetilde{\Gamma}(N)$ is generated by $\Gamma_1(N^2)$ and matrix $\mathbf{t}$, thus

$$\widetilde{\Gamma}(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \colon c \equiv 0 \pmod{N^2},\ a, d \equiv 1 \pmod{N} \right\}.$$

It is clear from the construction of the congruence subgroup $\widetilde{\Gamma}(N)$ that the quotient $\widetilde{\Gamma}(N)/\mathscr{H}^*$ defines a modular curve whose function field consists of all the functions on $X_1(N^2)$ invariant under t. ■

Remark. The congruence subgroup $\widetilde{\Gamma}(N)$ defined in Proposition 4.1.1 is the intersection of two other standard congruence subgroups $\Gamma_1(N)$ and $\Gamma_0(N^2)$.

As a direct consequence of Proposition 4.1.1, $X_1(N^2)/\langle t \rangle$ is a well-defined modular curve with a function field equal to

$$k(X_1(N^2)/\langle t \rangle) = \{ f \in k(X_1(N^2)) \colon f(t(E,R)) = f(E,R), \forall (E,R) \in S_1(N^2) \}.$$

The following proposition shows the relationship between congruence subgroups $\widetilde{\Gamma}(N)$ and $\Gamma_1(N^2)$.

Proposition 4.1.2. Let $\widetilde{\Gamma}(N)$ be a congruence subgroup defined as

$$\widetilde{\Gamma}(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \colon c \equiv 0 \pmod{N^2},\ a, d \equiv 1 \pmod{N} \right\}.$$

The congruence subgroup

$$\Gamma_1(N^2) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \colon \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right) \pmod{N^2} \right\}$$

is a normal subgroup of $\widetilde{\Gamma}(N)$ with index N.

Proof. The congruence subgroup $\widetilde{\Gamma}(N)$ is generated with the congruence subgroup $\Gamma_1(N^2)$ and matrix $\mathbf{t} = \left(\begin{smallmatrix} 1-N & -1 \\ N^2 & 1+N \end{smallmatrix}\right) \in \Gamma_1(N) \setminus \Gamma_1(N^2)$. To prove that $\Gamma_1(N^2)$ is a normal subgroup of $\widetilde{\Gamma}(N)$ it is enough to see that $\mathbf{t}^{-1} \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mathbf{t} \in \Gamma_1(N^2)$, for every matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(N^2)$. This is true because

$$\begin{aligned}
&\left(\begin{smallmatrix} 1-N & -1 \\ N^2 & 1+N \end{smallmatrix}\right) \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \left(\begin{smallmatrix} 1+N & 1 \\ -N^2 & 1-N \end{smallmatrix}\right) = \\
&= \left(\begin{smallmatrix} a(1-N)(1+N)-c(1+N)-N^2(b(1-N)-d) & a(1-N)+b(1-N)^2-c+d(1-N) \\ aN^2(1+N)+c(1+N)^2-N^2(bN^2+d(1+N)) & aN^2+bN^2(1-N)+c(1+N)+d(1-N)(1+N) \end{smallmatrix}\right) \\
&\equiv \left(\begin{smallmatrix} a(1-N)(1+N) & a(1-N)+b-2bN+d(1-N) \\ 0 & d(1-N)(1+N) \end{smallmatrix}\right) \pmod{N^2} \quad (\text{because } c \equiv 0 \pmod{N^2}) \\
&\equiv \left(\begin{smallmatrix} 1-N^2 & 2-2N+b-2bN \\ 0 & 1-N^2 \end{smallmatrix}\right) \pmod{N^2} \quad (\text{because } a,d \equiv 1 \pmod{N^2}) \\
&\equiv \left(\begin{smallmatrix} 1 & b-2bN-2N+2 \\ 0 & 1 \end{smallmatrix}\right) \pmod{N^2}.
\end{aligned}$$

To calculate the index of $\Gamma_1(N^2)$ in $\widetilde{\Gamma}(N)$ we will use the homomorphism $\pi_N \colon \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$, induced by reduction modulo N for $N \geq 1$. The kernel of π_N is the principal congruence subgroup $\Gamma(N)$, which is a normal subgroup of finite index in $\mathrm{SL}_2(\mathbb{Z})$. Any other congruence subgroup $\Gamma(N) \subset \widetilde{\Gamma}$ is of finite index in $\mathrm{SL}_2(\mathbb{Z})$ and it is a preimage of π_N, i.e. $\widetilde{\Gamma} = \pi_N^{-1}(\widehat{\Gamma})$ where $\widehat{\Gamma}$ is some subgroup of $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. The index $[\widetilde{\Gamma} \colon \Gamma(N)]$ is equal to $\#\widehat{\Gamma}$.

For $\widetilde{\Gamma}(N)$, after reducing elements of $\widetilde{\Gamma}(N)$ modulo N^2, the element c is equal to zero, for elements a and d we have $a,d \equiv 1 \pmod{N}$ and $a,b,c,d \in \mathbb{Z}/N^2\mathbb{Z}$. There are no additional conditions on b, but a and d must satisfy a condition for determinant: $ad \equiv 1 \pmod{N^2}$. Writing $a = 1+kN$ and $d = 1+lN$, where $k,l \in \{0,1,\ldots,N-1\}$, we get

$$(1+kN)(1+lN) = 1+N(k+l)+klN^2 \equiv 1 \pmod{N^2},$$

which implies $k+l \equiv 0 \pmod{N}$, so l depends completely on k. Therefore, d depends completely on a. Altogether, $\#\widehat{\widetilde{\Gamma}(N)} = N^3$. The index $[\widetilde{\Gamma}(N) \colon \Gamma(N^2)]$ is equal to

$$[\widetilde{\Gamma}(N) \colon \Gamma_1(N^2)][\Gamma_1(N^2) \colon \Gamma(N^2)],$$

thus

$$[\widetilde{\Gamma}(N) \colon \Gamma_1(N^2)] = \frac{\#\widehat{\widetilde{\Gamma}(N)}}{[\Gamma_1(N^2) \colon \Gamma(N^2)]} = \frac{\#\widehat{\widetilde{\Gamma}(N)}}{N^2} = \frac{N^3}{N^2} = N.$$

■

Remark. Some results regarding the modular curve $X(\widetilde{\Gamma}(N))$ can be directly derived from the properties of the diamond operator t, i.e. of the operator $\langle N+1\rangle$, and the relation of the congruence subgroup $\widetilde{\Gamma}(N)$ in regard to other standard congruence subgroups. Namely, congruence subgroups $\Gamma_1(N^2), \widetilde{\Gamma}(N)$, and $\Gamma_0(N^2)$ are related by the subgroup relation $\Gamma_1(N^2) \subset \widetilde{\Gamma}(N) \subset \Gamma_0(N^2)$, i.e. $X_1(N^2) \to X(\widetilde{\Gamma}(N)) \to X_1(N^2)$. Furthermore, the modular curve $X_0(N^2)$ is the quotient of the modular curve $X_1(N^2)$ with the operator $\langle N+1\rangle$ (as it is $X(\widetilde{\Gamma}(N))$). Then, the result of Proposition 4.1.2 is the consequence of the fact that the field extension $k(X_1(N^2))/k(X_0(N^2))$ is Galois and the order of the operator $\langle N+1\rangle$ is equal to N.

Remark. For the index $[\Gamma_1(N)\colon \Gamma_1(N^2)]$, first note that

$$[\mathrm{SL}_2(\mathbb{Z})\colon \Gamma_1(N^2)] = [\mathrm{SL}_2(\mathbb{Z})\colon \Gamma_1(N)][\Gamma_1(N)\colon \Gamma_1(N^2)],$$

so

$$[\Gamma_1(N)\colon \Gamma_1(N^2)] = \frac{[\mathrm{SL}_2(\mathbb{Z})\colon \Gamma_1(N^2)]}{[\mathrm{SL}_2(\mathbb{Z})\colon \Gamma_1(N)]}.$$

Using the known indices between congruence subgroups, indicated in Section 1.3, and the definition of the Euler totient function, we have the following equalities:

$$\begin{aligned}
[\Gamma_1(N)\colon \Gamma_1(N^2)] &= \frac{[\mathrm{SL}_2(\mathbb{Z})\colon \Gamma_1(N^2)]}{[\mathrm{SL}_2(\mathbb{Z})\colon \Gamma_1(N)]} \\
&= \frac{[\mathrm{SL}_2(\mathbb{Z})\colon \Gamma_0(N^2)][\Gamma_0(N^2)\colon \Gamma_1(N^2)]}{[\mathrm{SL}_2(\mathbb{Z})\colon \Gamma_0(N)][\Gamma_0(N)\colon \Gamma_1(N)]} \\
&= \frac{N^2 \prod\limits_{p|N^2}\left(1+\frac{1}{p}\right)\phi(N^2)}{N\prod\limits_{p|N}\left(1+\frac{1}{p}\right)\phi(N)} \\
&= \frac{N^2 \prod\limits_{p|N^2}\left(1+\frac{1}{p}\right) N^2 \prod\limits_{p|N^2}\left(1-\frac{1}{p}\right)}{N\prod\limits_{p|N}\left(1+\frac{1}{p}\right) N\prod\limits_{p|N}\left(1-\frac{1}{p}\right)} \\
&= N^2,
\end{aligned}$$

where we can cancel the products because if $p|N^2$ it also divides N.

Let $k(X_1(N^2))$ denote the function field corresponding to the modular curve $X_1(N^2)$.

Using the result of Proposition 4.1.2, the quotient $\widetilde{\Gamma}(N)/\Gamma_1(N^2)$ acts as a group of automorphism of $k(X_1(N^2))$ with fixed field $k(X(\widetilde{\Gamma}(N)))$, i.e.

$$k(X(\widetilde{\Gamma}(N))) = k(X_1(N^2))^{\widetilde{\Gamma}(N)/\Gamma_1(N^2)}.$$

From this, we have an equality of function fields:

$$k(X(\widetilde{\Gamma}(N))) = k(X_1(N^2))^{\mathbf{t}},$$

thus

$$k(X(\widetilde{\Gamma}(N))) = k(X_1(N^2)/\langle t\rangle).$$

We have shown that the function $b \in \pi_{1,N}^*(k(X_1(N)))$ is invariant under the operator t. Therefore,

$$\pi_{1,N}^*(k(X_1(N))) \overset{N}{\subset} k(X(\widetilde{\Gamma}(N))) = k(X_1(N^2)/\langle t\rangle),$$

where the degree of the extension is equal to the index of the subgroup. Returning to the equality (4.5), the modular function $g\colon (E,R) \mapsto g_{N,P}(-R)$ is an element of the field $k(X_1(N^2)/\langle t\rangle)$ with property $g^N = f$. The polynomial $x^N - f$ is a polynomial of degree N in $\pi_{1,N}^*(k(X_1(N)))[x]$ having g as a root. The equality (4.5) depends on the irreducibility of the polynomial $x^N - f$.

Lemma 4.1.3. Let f be a function defined on the set $S_1(N)$ with $(E,P) \mapsto f_{N,P}(-P)$, where $f_{N,P}$ is a normalized Miller function. Let g be a function defined on the set $S_1(N^2)$ with $(E,R) \mapsto g_{N,P}(-R)$, where $P = [N]R$ and $f_{N,P} \circ [N] = g_{N,P}^N$. Let $t^* \in \mathrm{Aut}(k(X_1(N^2)/X_1(N))$ be a pullback operator of the map

$$t(E,R) = (E,[N+1]R),\ (E,R) \in S_1(N^2).$$

Let $\pi_{1,N}^*\colon k(X_1(N)) \to k(X_1(N^2)$ be a pullback operator of the map

$$\pi_{1,N}((E,R)) = (E,[N]R).$$

The polynomial $x^N - f$ is an irreducible polynomial in $\pi_{1,N}^*(k(X_1(N)))[x]$.

Proof. The proof is given in Section 4.2.2. ■

The irreducibility of the polynomial $x^N - f$, as stated in Lemma 4.1.3, implies

$$\pi_{1,N}^*(k(X_1(N)))(g) = k(X_1(N^2)/\langle t \rangle),$$

which implies that b' is an element of $\pi_{1,N}^*(k(X_1(N)))(g)$. Therefore, the equality (4.5) holds, and it is possible to generalize radical isogenies using modular functions.

Example 4.1.4. Let $N = 5$ and E be an elliptic curve over the field

$$\mathbb{Q}_5(b,c) = \text{Frac}\, \frac{\mathbb{Q}[b,c]}{(F_5(b,c))}.$$

The Tate normal form for E, together with the point P of order 5 is

$$E\colon y^2 + (1-b)xy - by = x^3 - bx^2,\ P = (0,0). \tag{4.6}$$

In the case of $N = 5$ we have $F_5(b,c) = b - c = 0$, which implies a simpler Tate normal form (4.6). Having only a parameter b results in only one modular function $\mathbf{b}$ in $k(X_1(5))$. On the other side, the curve E' and the point P' of order 5 are given by

$$E' = E/\langle P \rangle\colon y^2 + (1-b')xy - b'y = x^3 - b'x^2,\ P' = (0,0).$$

For a point P, let R be a point of order 25 such that $[5]R = P$. The pair (E,R) is an enhanced elliptic curve for $\Gamma_1(25)$. The pullbacks $\pi_{1,5}^*, \pi_{2,5}^*$ and maps b, b' are defined as before.

From the example in [17, Section 4.], when $N = 5$, $f_{5,P}(-P) = b \in \mathbb{Q}_5(b)$. The fifth root of b is a function on $X_1(25)$, as $(E,R) \overset{g}{\mapsto} g_{5,[5]R}(-R)$ is a well-defined map with the property $g^5 = b$.

Observing the preimages of $\pi_{2,5}$, points $(E,R), (E,R+[1\cdot 5]R), (E,R+[2\cdot 5]R), (E,R+[3\cdot 5]R)$ and $(E,R+[4\cdot 5]R)$ are all mapped to the same final point. The operator t defined as $t(E,R) \mapsto (E,[5+1]R) = (E,[6]R)$ is of order 5 and $\langle t \rangle$ is isomorphic to $\mathbb{Z}/5\mathbb{Z}$. The congruence subgroup generated by $\Gamma_1(25)$ and matrix $\mathbf{t} = \left(\begin{smallmatrix} -4 & -1 \\ 25 & 6 \end{smallmatrix}\right)$ is

$$\widetilde{\Gamma}(5) = \left\{ \left(\begin{smallmatrix} \tilde{a} & \tilde{b} \\ \tilde{c} & \tilde{d} \end{smallmatrix}\right) \in \text{SL}_2(\mathbb{Z})\colon \left(\begin{smallmatrix} \tilde{a} & \tilde{b} \\ \tilde{c} & \tilde{d} \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right) \pmod{25} \right\}.$$

The functions b, b', g and every pullback by $\pi_{1,5}$ or $\pi_{2,5}$ are invariant under t, so they are also defined on the quotient $X_1(25)/\langle t \rangle$. For the number of elements in the group $\#\widehat{\widetilde{\Gamma}(5)}$, after reducing elements of $\widetilde{\Gamma}(5)$ modulo 25, the conditions on elements are $\tilde{c} =$

$0, \tilde{a}, \tilde{d} \equiv 1 \pmod 5$ and $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d} \in \mathbb{Z}/25\mathbb{Z}$. The value for $\tilde{a}$ and $\tilde{d}$ each can only be from the set $\{1, 6, 11, 16, 21\}$. Since the determinant of the matrix has to be 1 in $\mathrm{SL}(\mathbb{Z}/25\mathbb{Z})$, there are 25 possibilities for $\tilde{b}$. Therefore, there are 125 elements in this group, and the index $[\widetilde{\Gamma}(5)\colon \Gamma_1(25)]$ is equal to 5. The field extension $\pi_{1,5}^*(k(X_1(5))) \subset k(X_1(25)/\langle t \rangle)$ has degree 5, the polynomial $X^5 - b$ is irreducible in $\pi_{1,5}^*(k(X_1(5)))[x]$, has a well-defined root, thus

$$\pi_{1,5}^*(k(X_1(5)))(\sqrt[5]{b}) = k(X_1(25)/\langle t \rangle),$$

meaning $b' \in \pi_{1,5}^*(k(X_1(5)))(\sqrt[5]{b})$ and b' is a rational expression of $\sqrt[5]{b}$.

4.2. FIELD OF DEFINITION OF AUTOMORPHISMS OF THE MODULAR CURVE $X(\widetilde{\Gamma}(N))$

At the start of this chapter, we assumed that ζ_N, a primitive N-th root of unity for a positive integer N, is an element of our underlying field k. This was implicitly used throughout the chapter, especially in the discussion regarding the field equality (4.5). In this section, we give a justification for that assumption. The necessary background was introduced in Section 1.3.2.

We will show that the automorphisms of the modular curve $X(\widetilde{\Gamma}(N))$ are defined over the field $\mathbb{Q}(\zeta_N)$, and consequently over the field k, where k is a field satisfying $\mathrm{char}(k) \nmid N$ and $\zeta_N \in k$. To accomplish this, we will utilize the proposition [34, Section 5, Proposition 5.14.] outlined in Section 4.2.2. This proposition gives a field of definition of automorphisms of modular curves using the subgroups of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ and a matrix in $\mathrm{SL}_2(\mathbb{Z})$ that normalizes those subgroups.

4.2.1. Automorphisms of a modular curve

Let k be a field, and let $\bar{k}$ denote its algebraic closure. We first give a short introduction to the automorphisms of modular curves.

Let Γ_1 and Γ_2 be two congruence subgroups such that Γ_1 is a normal subgroup of Γ_2. Let $X(\Gamma_1)$ and $X(\Gamma_2)$ be the associated modular curves. The congruence subgroup Γ_2 acts on the modular curve $X(\Gamma_1)$. First, using the fractional linear transformation, Γ_2 acts on the upper half-plane: for a matrix $\gamma_2 \in \Gamma_2$ and an element $z \in \mathscr{H}$ we have $z \mapsto \gamma_2 z$. This is also a definition of an automorphism of the upper half-plane. Second, if we extend that action, the congruence subgroup Γ_2 will act on the modular curve $X(\Gamma_1)$ by the following: $\Gamma_1 z \mapsto \gamma_2(\Gamma_1 z)$, where $\Gamma_1 z$ is an element of the modular curve $X(\Gamma_1)$. The congruence subgroup Γ_1 will also (trivially) act on the modular curve $X(\Gamma_1)$, and therefore, we have the action of the quotient group Γ_2/Γ_1 on the same modular curve. The quotient group Γ_2/Γ_1 consists of all the left cosets of Γ_1 in Γ_2, i.e. for a matrix $\gamma_2 \in \Gamma_2$, we have the set

$$\gamma_2\Gamma_1 = \{\gamma_2\gamma_1 : \gamma_1 \in \Gamma_1\}.$$

Each coset in the quotient group Γ_2/Γ_1 corresponds to an automorphism of the modular curve $X(\Gamma_1)$. Let g_2 be a matrix from the coset $\gamma_2\Gamma_1$. There exist $g_1 \in \Gamma_1$ such that $g_2 = \gamma_2 g_1$. Now, $g_2\Gamma_1 g_2^{-1} = \gamma_2 g_1 \Gamma_1 g_1^{-1}\gamma_2^{-1}$, and because $g_1\Gamma_1 g_1^{-1}$ is a subset of Γ_1, and $\gamma_2\Gamma_1\gamma_2^{-1} \in \Gamma_1$ because Γ_1 is normal subgroup of Γ_2, we have $g_2\Gamma_1 g_2^{-1} = \Gamma_1$, so the automorphism is well-defined as it does not depend on the coset representative. Automorphisms constructed in this way induce automorphisms of the field $\overline{k}(X(\Gamma_1))/\overline{k}(X(\Gamma_2))$. Furthermore, this is a Galois extension and $\mathrm{Gal}(\overline{k}(X(\Gamma_1))/\overline{k}(X(\Gamma_2))) \simeq \Gamma_2/\Gamma_1$. In other words, the quotient Γ_2/Γ_1 acts as a group of automorphisms of $\overline{k}(X(\Gamma_1))$ with fixed field $\overline{k}(X(\Gamma_2))$.

The modular curve $X(\widetilde{\Gamma}(N))$ is defined as a quotient $X_1(N^2)/\langle t\rangle$, where t (or more precisely t_{N+1}, see Section 4.1.1) is a restriction of an automorphism of the modular curve $X_1(N^2)$. Because of this, in order to describe automorphisms of $X(\widetilde{\Gamma}(N))$, the first idea is to look at the automorphisms of modular curve $X_1(N^2)$. Because of the relation $\Gamma_1(N^2) \subset \widetilde{\Gamma}(N) \subset \Gamma_1(N)$, the obvious candidate for the automorphism group is $\Gamma_1(N)/\Gamma_1(N^2)$, but this is not a good path because $\Gamma_1(N^2)$ is not a normal subgroup of $\Gamma_1(N)$. So, we need to go a step further, i.e. find a congruence subgroup Γ such that $\Gamma \subset \Gamma_1(N^2) \subset \widetilde{\Gamma}(N) \subset \Gamma_1(N)$ and Γ is a normal subgroup of $\Gamma_1(N)$. Then, the quotient $\Gamma_1(N)/\Gamma$ will act as a group of automorphisms of $\overline{k}(X(\Gamma))$ with fixed field $\overline{k}(X_1(N))$, and the field extension $\overline{k}(X(\Gamma))/\overline{k}(X_1(N))$ will be a Galois extension. The automorphism t_{N+1} should be a restriction of some automorphism of $X(\Gamma) \mapsto X(\Gamma_1(N))$.

Let m,n be positive integers such that $m|n$, and let $\Gamma(m,n)$ denote the congruence subgroup

$$\Gamma(m,n) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \colon a,d \equiv 1 \pmod{n},\ c \equiv 0 \pmod{n},\ b \equiv 0 \pmod{m} \right\}.$$

This is indeed a congruence subgroup because $\Gamma(n) \subset \Gamma(m,n)$. The associated modular curve is a quotient $\Gamma(m,n)/\mathscr{H}$, denoted by $Y(m,n)$ that parametrizes triples (E,P_m,P_n) where E is an elliptic curve, P_m is a point of order m and P_n is a point of order n. Additionally, P_m and P_n are independent points (so they generate a subgroup isomorphic to $\mathbb{Z}/m\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z}$), and there is a condition on Weil pairing such that $e_m(P_m, [\frac{n}{m}]P_n) = \zeta_m$, where ζ_m is fixed m-th root of unity. Let $X(m,n)$ be the compactification of $Y(m,n)$.

If we let $m = N$ and $n = N^2$, the congruence subgroup is:

$$\Gamma(N,N^2) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \colon a,d \equiv 1 \pmod{N^2},\ c \equiv 0 \pmod{N^2},\ b \equiv 0 \pmod{N} \right\}. \tag{4.7}$$

The subgroup $\Gamma(N,N^2)$ is a normal subgroup of $\Gamma_1(N)$. The verification is straightforward, let $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(N)$ and let $\left(\begin{smallmatrix} x & y \\ w & z \end{smallmatrix}\right) \in \Gamma(N,N^2)$. The product is equal to:

$$\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} x & y \\ w & z \end{smallmatrix}\right)\left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right) = \begin{pmatrix} d(ax+bw)-c(ay+bz) & a(ay+bz)-b(ax+bw) \\ d(cx+dw)-c(cy+dz) & a(cy+dz)-b(cx+dw) \end{pmatrix},$$

and it should be an element of $\Gamma(N,N^2)$. For example, the element $d(ax+bw)-c(ay+bz)$ should be congruent to one modulo N^2. We have the following:

$$\begin{aligned} d(ax+bw)-c(ay+bz) &\equiv da+dbw-cay-cb \pmod{N^2} \quad (\text{because } x,z \equiv 1 \pmod{N^2}) \\ &\equiv da-cay-cb \pmod{N^2} \quad (\text{because } w \equiv 0 \pmod{N^2}) \\ &\equiv 1-cay \pmod{N^2} \quad (\text{because } ad-cb \text{ is the determinant}). \end{aligned}$$

This implies that we should have $1-cay \equiv 1 \pmod{N^2}$, i.e. $cay \equiv 0 \pmod{N^2}$. The elements c and y are both divisible with N, so there exists an integer k such that $c = k \cdot N$ and an integer l such that $y = l \cdot N$. Therefore the product cay is equal to $klaN^2$ which is congruent to zero modulo N^2 as required. The modular condition of other elements is checked in a similar way.

We can define the quotient $\Gamma_1(N)/\Gamma(N,N^2)$, and it will act as a group of automorphisms on $\overline{k}(X(N,N^2))$ with fixed field $\overline{k}(X_1(N))$. The field extension $\overline{k}(X_1(N)) \subset \overline{k}(X(N,N^2))$ is a Galois extension with a degree equal to N^3. Figure 4.2 shows the relation between function fields.

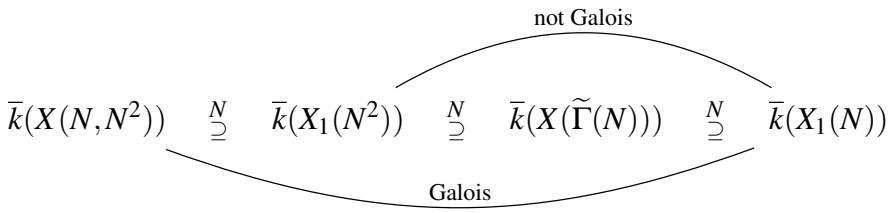

Figure 4.2: Function fields related to congruence subgroup (4.7)

The Galois group of the field extension $\overline{k}(X(N,N^2))/\overline{k}(X_1(N))$ is generated by the

following three automorphisms:

$$\alpha\colon (E,P,Q) \mapsto (E,P+[N]Q,Q), \tag{4.8a}$$

$$t\colon (E,P,Q) \mapsto (E,P,[N+1]Q), \tag{4.8b}$$

$$\omega\colon (E,P,Q) \mapsto (E,P,P+Q), \tag{4.8c}$$

where P is a point of order N, Q is a point of order N^2, they are independent and such that $e_N(P,[N]Q) = \zeta_N$.

The notation for the automorphism t coincides with the notation for the automorphism t, i.e. t_{N+1} from Section 4.1.1. We recall that the subgroup generated by t_{N+1} was used to define the modular curve $X(\widetilde{\Gamma}(N)) = X_1(N^2)/\langle t_{N+1}\rangle$. Keeping the same notation is the consequence of the fact that the automorphism t_{N+1} is a restriction of the automorphism (4.8b) on $X(N,N^2) \mapsto X_1(N^2)$. This is represented by Figure 4.3.

$$\begin{array}{ccc} X(N,N^2) \ni (E,P,Q) & \xmapsto{\ t\ } & (E,P,[N+1]Q) \\ \downarrow & & \downarrow \\ X_1(N^2) \ni (E,Q) & \xmapsto{\ t_{N+1}\ } & (E,[N+1]Q) \end{array}$$

Figure 4.3: Automorphism t

Furthermore, the quotient of the modular curve $X(N,N^2)$ and the group generated by the automorphism α is the modular curve $X_1(N^2)$. This is obvious because α identifies the points of order N^2, and those points are the only ones considered in the map $(E,P,Q) \mapsto (E,Q)$. We can conclude that the quotient of the modular curve $X(N,N^2)$ and of the group generated by the automorphisms t and α is precisely $X(\widetilde{\Gamma}(N))$, i.e.

$$\overline{k}(X(\widetilde{\Gamma}(N))) = (\overline{k}(X(N,N^2))^{\langle\alpha\rangle})^{\langle t\rangle}.$$

We still need to discuss the automorphism (4.8c). We are interested in the restriction of this automorphism to the modular curve $X(\widetilde{\Gamma}(N))$. This restriction will generate the Galois group of the extension $\overline{k}(X(\widetilde{\Gamma}(N)))/\overline{k}(X_1(N))$, where we note that $\widetilde{\Gamma}(N)$ is a normal subgroup of $\Gamma_1(N)$. The definition of the operator t_{N+1} from the Section 4.1.1 was influenced by preimages of the pullback operators, i.e. we have seen that, for a point R of order N^2, points $(E,R),(E,R+[1\cdot N]R),\ldots,(E,R+[(N-1)\cdot N]R)$ are all mapped to the same point. The orbit of the point R is the set $\{R+[k]([N]R),\ k=1,\ldots,N-1\}$.

$$\begin{array}{ccc} X(N,N^2) \ni (E,P,Q) & \xmapsto{\omega} & (E,P,P+Q) \\ \downarrow & & \downarrow \\ X_1(N^2) \ni (E,Q) & & (E,P+Q) \\ \downarrow & & \downarrow \\ X(\widetilde{\Gamma}(N)) \ni (E,Q+\langle t\rangle) & \xmapsto{\omega} & (E,P+Q+\langle t\rangle) \end{array}$$

Figure 4.4: Automorphism ω

We need to check that the restriction of ω on $X(\widetilde{\Gamma}(N))$ (we are keeping the same notation for the restriction) is a well-defined map. It is enough to check that for a fixed point Q, points $(E,P+Q+\langle t\rangle)$ and $(E,P'+Q+\langle t\rangle)$ represent the same element in $X(\widetilde{\Gamma}(N))$, for $(E,P,Q),(E,P',Q) \in X(N,N^2)$. To put it another way, points $(E,P+Q)$ and $(E,P'+Q)$ in $X_1(N^2)$ should be elements of the same $\langle t\rangle$ orbit. Elements of the orbit of the point $P+Q$ can be written as $P+Q+[k]([N]Q)$, for $k=1,\ldots,N-1$. The point $P'+Q$ will be an element of that orbit if $P'+Q=P+Q+[k]([N]Q)$, i.e. if $P'=P+[k]([N]Q)$. This follows from the definition of the modular curve $X(N,N^2)$, since $e_N(P',[N]Q) = e_N(P,[N]Q)=\zeta_N$, the bilinearity of Weil pairing implies $e_N(P-P',[N]Q)=1$, and because $e_N([N]Q,[N]Q)=1$, the point $P-P'$ is an element of a group $\langle [N]Q\rangle$.

4.2.2. Field of definition of automorphisms

In Section 1.3.2, we introduced the modular curve X_H for a subgroup $H \in \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$. In the preceding section we introduced the modular curve $X(N,N^2)$ along with the automorphisms α, t and ω. The goal of this section is to determine the field over which these automorphisms are defined. To achieve this, we will utilize the subgroups of $\mathrm{GL}_2(\mathbb{Z}/N^2\mathbb{Z})$ associated with (concerning the reduction modulo N^2 map) the modular curve $X(N,N^2)$.

We denote the quotient $X_H/\mathscr{H}^*$ by $X_H(\mathbb{C})$. Furthermore, we denote the regular 2×2 matrices with elements from $\mathbb{Q}$ and with positive determinant by $\mathrm{GL}_2^+(\mathbb{Q})$, and by $\mathrm{PGL}_2^+(\mathbb{Q}) = \mathrm{GL}_2^+(\mathbb{Q})/\{\pm I\}$ the projective linear group. The map $\mathbb{P}\colon \mathrm{GL}_2^+(\mathbb{Q}) \to \mathrm{PGL}_2^+(\mathbb{Q})$ is the natural map. The following is the definition of a modular automorphism of X_H, see [34, Definition 5.1.].

Definition 4.2.1. Let H be a subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ such that $\det(H) = (\mathbb{Z}/N\mathbb{Z})^\times$.

An automorphism of X_H defined over $\mathbb{C}$ is modular if its action on $X_H(\mathbb{C})$ is described by a fractional linear transformation associated with a matrix in $\mathrm{PGL}_2^+(\mathbb{Q})$ normalizing $\mathbb{P}(\Gamma_H)$.

How to find an automorphism and its field of definition is covered in the following proposition.

Proposition 4.2.2. Let N be a positive integer, let H be a subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ such that $\det(H) = (\mathbb{Z}/N\mathbb{Z})^\times$, and let $H' = H \cap \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. Let $M \in \mathrm{SL}_2(\mathbb{Z})$ be a matrix such that its reduction $(M \bmod N)$ normalizes H'. Then M defines a modular automorphism of X_H which is defined over the cyclotomic field $\mathbb{Q}(\zeta_N)$. Moreover, this automorphism is defined over $\mathbb{Q}$ if and only if $(M \bmod N)$ normalizes H.

Proof. See Proposition [34, Section 5, Proposition 5.14.]. ■

Using this proposition, our goal is to determine the field over which the automorphisms of the modular curve $X(N,N^2)$ are defined. This field will also be the field of definition of automorphisms of the modular curve $X(\widetilde{\Gamma}(N))$, as every automorphism there can be regarded as a restriction of an automorphism of $X(N,N^2)$. First, we need to identify subgroups of $\mathrm{GL}_2(\mathbb{Z}/N^2\mathbb{Z})$ corresponding to the congruence subgroup $\Gamma(N,N^2)$ defined by (4.7). These subgroups should be related to $\Gamma(N,N^2)$ in the sense that their preimage under the reduction modulo N^2 map is exactly that subgroup.

We start with the subgroup H. From Proposition 4.2.2, we know that the matrices from this subgroup have the determinant equal to $\det(H) = (\mathbb{Z}/N\mathbb{Z})^\times$. So, the subgroup H is equal to:

$$H := \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Z}/N^2\mathbb{Z}) \colon c \equiv 0 \pmod{N^2},\ d \equiv 1 \pmod{N^2},\ b \equiv 0 \pmod{N} \right\},$$

i.e. we have eliminated the condition on the first element. Furthermore, again according to Proposition 4.2.2, the second subgroup H' is defined as the intersection $H' := H \cap \mathrm{SL}_2(\mathbb{Z}/N^2\mathbb{Z})$. Additionally, we let H'' be the subgroup between them, i.e.

$$H'' := \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Z}/N^2\mathbb{Z}) \colon c \equiv 0 \pmod{N^2},\ d \equiv 1 \pmod{N^2},\ a \equiv 1 \pmod{N},\ b \equiv 0 \pmod{N} \right\}.$$

The relation between the subgroups is $H' \subset H'' \subset H$.

Similar to the matrix $\mathbf{t}$ from Proposition 4.1.1, we let $\omega = \left(\begin{smallmatrix} 1-N & -N \\ N & 1+N \end{smallmatrix}\right) \in \Gamma_1(N) \setminus \Gamma_1(N^2)$ be a matrix representation of the automorphism ω. If ω normalizes H', it will define a modular automorphism of X_H over a cyclotomic field $\mathbb{Q}(\zeta_{N^2})$.

Let $h' = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be a matrix from the subgroup H'. Multiplying h' with ω gives us:

$$\begin{aligned}\omega \cdot h' \cdot \omega^{-1} &= \left(\begin{smallmatrix} 1-N & -N \\ N & 1+N \end{smallmatrix}\right)\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} 1+N & N \\ -N & 1-N \end{smallmatrix}\right) \\ &= \begin{pmatrix} (1+N)(a(1-N)-cN)-N(b(1-N)-dN) & N(a(1-N)-cN)+(1-N)(b(1-N)-dN) \\ (1+N)(aN+c(N+1))-N(bN+d(1+N)) & N(aN+c(1+N))+(1-N)(bN+d(1+N)) \end{pmatrix},\end{aligned}$$

which is an element of H' because:

- For the determinant we have $\det(\omega \cdot h' \cdot \omega^{-1}) = \det(\omega)\det(h')\det(\omega^{-1}) = 1$ in $\mathbb{Z}/N^2\mathbb{Z}$, since h' is an element of $\mathrm{SL}_2(\mathbb{Z}/N^2\mathbb{Z})$. Furthermore, as $\det(h') = ad - cb = 1$ in $\mathbb{Z}/N^2\mathbb{Z}$, conditions on the elements b, c and d imply $a \equiv 1 \pmod{N^2}$.

- The condition on the second element is satisfied because:

$$\begin{aligned} N(a(1-N) - cN) + (1-N)(b(1-N) + dN) = \\ = -aN^2 + bN^2 - cN^2 + dN^2 + aN - bN - bN - dN + b \\ = b \equiv 0 \pmod{N}.\end{aligned}$$

- The condition on the third element is satisfied because:

$$\begin{aligned}(1+N)(aN + c(N+1)) - N(bN + d(1+N)) = \\ = aN^2 - bN^2 + cN^2 - dN^2 + aN + cN + cN - dN + c \\ = (a-d)N \equiv 0 \pmod{N^2},\end{aligned}$$

  as $a, d \equiv 1 \pmod{N^2}$ implies $a - d \equiv 0 \pmod{N^2}$.

- And the condition on the fourth element is satisfied because:

$$\begin{aligned} N(aN + c(1+N)) + (1-N)(bN + d(1+N)) = \\ = aN^2 - bN^2 + cN^2 - dN^2 + bN + cN + dN - dN + d \\ \equiv bN + 1 \pmod{N^2},\end{aligned}$$

  so we need to have $bN + 1 \equiv 1 \pmod{N^2} \Leftrightarrow bN \equiv 0 \pmod{N^2}$, which is true because $b \equiv 0 \pmod{N}$.

According to Proposition 4.2.2, the matrix ω defines a modular automorphism of X_H over the field $\mathbb{Q}(\zeta_{N^2})$. Additionally, as ω does not normalize the subgroup H, this automorphism is not defined over $\mathbb{Q}$.

We can go a step further. Since the matrix ω also normalizes the congruence subgroup H'' (the verification is similar to that of the subgroup H'), this implies a possibility of a different field of definition. We can gain more insights by examining the determinants of the matrices H', H'', and H. For H, as there is no condition on its element a, determinant $\det(H)$ is equal to $(\mathbb{Z}/N^2\mathbb{Z})^\times$. Regarding H'', from the conditions on elements a and d, we find $(1+kN)(1+lN^2) \equiv 1+kN \pmod{N^2}$, so the determinant is equal to $\det(H'') = \{1+kN \in \mathbb{Z}/N^2\mathbb{Z},\ k \in \mathbb{Z}\}$, i.e. the determinant is the set of all elements whose modulo N is equal to 1. For the subgroup H', as it is an intersection with $\mathrm{SL}_2(\mathbb{Z}/N^2\mathbb{Z})$, the determinant is equal to $\{1\}$.

Using the action of the determinant of the subgroup H'', we will show the equality $\mathbb{Q}(\zeta_{N^2})^{\det(H'')} = \mathbb{Q}(\zeta_N)$, i.e. we will show that the field of definition of the automorphisms is the field $\mathbb{Q}_N(\zeta_N)$.

Let ζ_{N^2} be N^2-th primitive root of unity:

- $\zeta_{N^2}^{N^2} = 1$, and
- $\zeta_{N^2}^{x} \neq 1$, for positive integer $x < N^2$.

Consider an element $1+kN \in \mathbb{Z}/N^2\mathbb{Z}$, where k is a positive integer. Let m be an integer such that $\zeta_{N^2}^m$ is fixed by $1+kN$, for every k. Put differently, element $\zeta_{N^2}^m$ is fixed by the determinant $\det(H'')$. We have the following equalities:

$$\begin{aligned}\left(\zeta_{N^2}^m\right)^{1+kN} &= \zeta_{N^2}^m,\\ \zeta_{N^2}^{m+mkN} &= \zeta_{N^2}^m,\\ \zeta_{N^2}^{mkN} &= 1.\end{aligned}$$

The property of a primitive root, where for a n-th primitive root z, we have

$$z^a = z^b \Leftrightarrow a \equiv b \pmod{n},$$

implies:

$$mkN \equiv 0 \pmod{N^2} \Rightarrow mk \equiv 0 \pmod{N},$$

therefore $N|mk$ for every k, implying $N|m$. Let l be an integer such that $Nl = m$. Then,

$$\zeta_{N^2}^m = \zeta_{N^2}^{Nl} = \zeta_N^l,$$

where the last equality follows from another property of primitive root: if z is n-th primitive root, the power z^x is an a-th primitive root for $a = \frac{n}{\gcd(n,x)}$. Altogether,

$$\left(\zeta_{N^2}^m\right)^{1+kN} = \zeta_{N^2}^m = \zeta_N^l.$$

Remark. A similar approach can be applied directly to the congruence subgroup

$$\widetilde{\Gamma}(N) \coloneqq \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \colon c \equiv 0 \pmod{N^2},\ a,d \equiv 1 \pmod{N} \right\}.$$

In this case the three subgroups are:

$$H \coloneqq \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Z}/N^2\mathbb{Z}) \colon c \equiv 0 \pmod{N^2},\ d \equiv 1 \pmod{N} \right\},$$

$$H'' \coloneqq \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Z}/N^2\mathbb{Z}) \colon c \equiv 0 \pmod{N^2},\ a,d \equiv 1 \pmod{N} \right\},$$

and the third is defined as the intersection $H' \coloneqq H \cap \mathrm{SL}_2(\mathbb{Z}/N^2\mathbb{Z})$.

In conclusion, the fixed field under the action of the determinant $\det(H'')$ on elements from the field $\mathbb{Q}(\zeta_{N^2})$ is equal to $\mathbb{Q}(\zeta_N)$. Consequently, the matrix ω defines (modular) automorphisms of X_H over the cyclotomic field $\mathbb{Q}(\zeta_N)$. All automorphisms of $X(N,N^2)$ (α, ω and t), and thus their restrictions, are well-defined on a field $k(\zeta_N)$, when $\mathrm{char}(k) \nmid N$.

Let, as before, k be a field, $N \geq 4$ an integer such that $\mathrm{char}(k) \nmid N$, and $\zeta_N \in k$. The field k is the field of definition of automorphisms of modular curve $X(\widetilde{\Gamma}(N))$. To conclude this section, we provide the proof of Lemma 4.1.3:

Proof. The function g is invariant under the operator t, so it is defined on the quotient $X(\widetilde{\Gamma}(N))$. We will show that the function g is modular only for that congruence subgroup, so g cannot be an element of some other field $k(X(\Gamma))$ with $\widetilde{\Gamma}(N) \subsetneq \Gamma \subset \Gamma_1(N)$. This implies that the field extension $\pi_{1,N}^*(k(X_1(N)))(g)$ has degree N over $k(X_1(N))$.

Let P be a point of order N, and let R and R' be points of order N^2 such that $[N]R = P$ and $[N]R' = P$. Additionally, we assume that the function g has the same values in those points, i.e. $g(E,R) = g(E,R')$. Using the definition of the function g, we have

$$g_{N,[N]R}(-R) = g_{N,[N]R'}(-R') = g_{N,[N]R}(-R').$$

The previously defined Weil pairing implies:

$$1 = \frac{g_{N,[N]R}(-R')}{g_{N,[N]R}(-R)} = \frac{g_{N,P}((R-R')-R)}{g_{N,P}(-R)} = e_N(P, R-R').$$

The point $R-R'$ belongs to $E[N]$ because, by assuming $[N]R' = [N]R = P$, we have $[N](R-R') = P-P = \mathcal{O}_E$. Therefore, $e_N(P, R-R')$ is consistent with the definition of the Weil pairing.

Let E be a fixed elliptic curve, P a point of order N on that curve, and R a point of order N^2 on that curve such that $P = [N]R$. From the discussion before, for every R, we have $e_N(P, R-R') = 1$, and since the Weil pairing is nondegenerate, the point $R-R'$ belongs to the subgroup $\langle P \rangle$. As a consequence, the point R' can be written as $R + [l]P$ for some $l \in \mathbb{Z}$, which depends on R.

In comparison to the operator t, since g is invariant under t, we have:

$$g(E,R) = g(t(E,R)) = g(E, [N+1]R) = g(E, R+P),$$

which implies $g(E,R) = g(E, R+[k]P)$ for every $k \in \mathbb{Z}$.

In conclusion, if the function g has the same values for any two different points of order N^2, then those two points are dependent on each other, and moreover, the relation between them directly follows from the invariant property of the function g under the operator t.

In the discussion before, we have seen that ω is a generator of the Galois group $\mathrm{Gal}(k(X_1(N))/k(X(\widetilde{\Gamma}(N))))$. If g is invariant under some other operator, i.e. if g is defined on some quotient of $X(\widetilde{\Gamma}(N))$, this would imply that g should be invariant under ω^k, for some positive integer k, such that ω^k is not the identity. This would imply that there are some points $\overline{R}$ and $\overline{R'}$ such that $\overline{R} - \overline{R'} \neq [\bar{l}]P$, for any integer $\bar{l}$. But this is in contradiction with the invariance property of g under the operator t.

In other words, g is modular only for the congruence subgroup $\widetilde{\Gamma}(N)$. This implies that the function field $\pi_{1,N}^*(k(X_1(N)))(g)$ is an extension of the degree exactly N over $k(X_1(N))$. The roots of the polynomial $x^N - f$ are of the form $\zeta_N^n g$, where ζ_N represents the N-th root of unity and n is a positive integer. If we assume that this polynomial is not irreducible, then we could find two nonconstant polynomials $f_1, f_2 \in k(X_1(N))[x]$, such that $x^N - f = f_1(x) f_2(x)$. However, this would lead to a contradiction since g is a root for

f_1 and has a degree greater than or equal to N, which is the degree of g. Therefore, the polynomial $x^N - f$ is irreducible. ■

5. Radical isogenies in the language of modular curves - the case $X_0(N)$

The background set in the previous chapter, including the results on radical isogenies, offer a way to address the open problem introduced in Example 3.2.5. In this chapter, we will delve deeper into the concepts discussed previously, now extending our focus to include the modular curve $X_0(N)$. Throughout this chapter, we will maintain a consistent use of notations and assumptions introduced in Chapter 4.

5.1. Extending to $X_0(N)$

Let k be a field, let $N \geq 5$ be a positive integer such that $\mathrm{char}(k) \nmid N$, and let $\overline{k}$ denote the algebraic closure of k. Let β be a function on enhanced elliptic curves for $\Gamma_0(N)$, i.e. an element of $\overline{k}(X_0(N))$. For example, we can take β to be a Hauptmodul[1] for $\overline{k}(X_0(N))$. Such Hauptmodul will exist if the genus of the modular curve is zero. We let the pullback operators $\pi_{1,N}^*$ and $\pi_{2,N}^*$ to be defined by (4.4), but now on the algebraic closure $\overline{k}$, and ψ_N^* is the pullback operator

$$\psi_N^* \colon \overline{k}(X_0(N)) \to \overline{k}(X_1(N)), \text{ of the map } \psi_N((E,P)) = (E, \langle P \rangle),$$

where (E,P) is an enhanced elliptic curve for $X_1(N)$. Applying the compositions $\pi_{1,N}^* \circ \psi_N^*$ and $\pi_{2,N}^* \circ \psi_N^*$ to the functions from $\overline{k}(X_0(N))$ results in elements of $\overline{k}(X_1(N^2))$. From

[1] A Hauptmodul for a congruence subgroup Γ is a function that generates the field of modular functions for Γ.

now on, we will identify the function β with $\beta := \pi_{1,N}^*(\psi_N^*(\beta))$ and define $\beta' := \pi_{2,N}^*(\psi_N^*(\beta))$. Both β and β' are elements of $\overline{k}(X_1(N^2))$.

In the last chapter, we introduced the congruence subgroup $\Gamma(N,N^2)$ and its associated modular curve $X(N,N^2)$ to find the field of definition of automorphism of the modular curve $X(\widetilde{\Gamma}(N))$. Furthermore, the choice to introduce the congruence subgroup $\Gamma(N,N^2)$ was additionally motivated by the fact that the subgroup $\Gamma_1(N^2)$ was not a normal subgroup of $\Gamma_1(N)$. Consequently, we could not build automorphisms from their quotient as outlined in Section 4.2.1, and utilize parts of Galois theory on associated function fields. Resolving the problem introduced in Example 3.2.5 involves determining the function field to which the function β' belongs. Referring back to the previous chapter, the generalization of radical isogenies to modular curves became feasible due to the connection between the functions b and b' through the equality (4.5). Specifically, the function b' was an element of a field extension of the field $\pi_{1,N}^*(k(X_1(N)))$, which contains the function b as an element. In this chapter's extended setting, we aim to achieve a similar result. If radical isogeny formulas exist on $X_0(N)$, it should be possible to express β' as an element of a function field whose definition depends on the function β. Because the congruence subgroup $\Gamma_1(N^2)$ is not a normal subgroup of $\Gamma_0(N)$, to achieve this we will introduce another pullback that will enable us to realize the functions β and β' as elements of the function field $\overline{k}(X(N,N^2))$. In this way we will be able to use the automorphisms α, t and ω defined by (4.8), and the fact that they generate the Galois group of the field extension $\overline{k}(X(N,N^2))/\overline{k}(X_1(N))$.

Let E be an elliptic curve over the field k. Let R be a point of order N^2 on E. Let P and P' be points of order N such that $P = [N]R$ and $e_N(P', [N]R) = \zeta_N$. Let Φ_N be a pullback operator

$$\Phi_N^*\colon \overline{k}(X_1(N^2)) \to \overline{k}(X(N,N^2)), \text{ of the map } \Phi_N((E,P',R)) = (E,R),$$

where (E,P',R) is an enhanced elliptic curve for $X(N,N^2)$. Maps and functions are visualized in Figure 5.1.

The subgroup $\Gamma(N,N^2)$ is a normal subgroup of $\Gamma_0(N)$. The verification is straightforward: let $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$ and $\left(\begin{smallmatrix} x & y \\ w & z \end{smallmatrix}\right) \in \Gamma(N,N^2)$. The condition for a normal subgroup

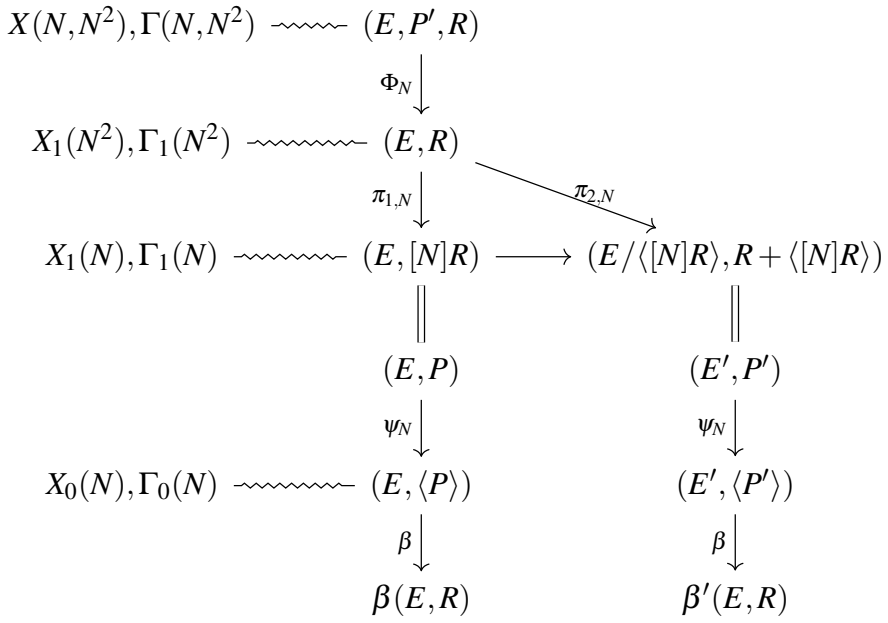

Figure 5.1: Maps between enhanced elliptic curves, including $X_0(N)$

implies that the product

$$\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} x & y \\ w & z \end{smallmatrix}\right)\left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right) = \begin{pmatrix} d(ax+bw)-c(ay+bz) & a(ay+bz)-b(ax+bw) \\ d(cx+dw)-c(cy+dz) & a(cy+dz)-b(cx+dw) \end{pmatrix},$$

should be an element of $\Gamma(N,N^2)$. Therefore, the element $a(cy+dz)-b(cx+dw)$ should be congruent to one modulo N^2. We have the following:

$$\begin{aligned} a(cy+dz)-b(cx+dw) &\equiv acy+ad-bc-bdw \pmod{N^2} \quad \text{(because } x,z \equiv 1 \pmod{N^2}\text{)} \\ &\equiv acy+ad-bc \pmod{N^2} \quad \text{(because } w \equiv 0 \pmod{N^2}\text{)} \\ &\equiv 1+acy \pmod{N^2} \quad \text{(because } ad-cb \text{ is the determinant).} \end{aligned}$$

Similar to before, this implies that we should have $1+acy \equiv 1 \pmod{N^2}$, i.e. $acy \equiv 0 \pmod{N^2}$. The elements c and y are both divisible with N, so there exists an integer k such that $c = k \cdot N$ and an integer l such that $y = l \cdot N$. Therefore the product acy is equal to $aklN^2$ which is congruent to zero modulo N^2 as required. The modular condition on the element $d(ax+bw)-c(ay+bz)$ is checked similarly. Furthermore, the element

$a(ay+bz)-b(ax+bw)$ should be congruent to zero modulo N. We have the following:

$$\begin{aligned} a(ay+bz)-b(ax+bw) &\equiv abz-bax-bbw \pmod N \quad (\text{because } y\equiv 0 \pmod N)\\ &\equiv abz-bax \pmod N \quad (\text{because } w\equiv 0 \pmod{N^2})\\ &\equiv abz-bax \pmod N \quad (\text{because } x,z\equiv 1 \pmod{N^2})\\ &\equiv 0 \pmod N. \end{aligned}$$

Finally, for the element $d(cx+dw)-c(cy+dz)$, which should be congruent to zero modulo N^2, we have the following:

$$\begin{aligned} d(cx+dw)-c(cy+dz) &\equiv dcx-ccy-cdz \pmod{N^2} \quad (\text{because } w\equiv 0 \pmod{N^2})\\ &\equiv dc-ccy-cd \pmod{N^2} \quad (\text{because } x,z\equiv 1 \pmod{N^2})\\ &\equiv -ccy\equiv 0 \pmod{N^2}, \end{aligned}$$

because $c,y\equiv 0 \pmod N$.

The Galois group of the field extension $\overline{k}(X(N,N^2))/\overline{k}(X_1(N))$ is generated by automorphisms α,t and ω. Let λ be an automorphism on $X(N,N^2)$ defined by

$$\lambda\colon (E,P',Q')\mapsto (E,[\lambda]P',[\lambda^{-1}]Q'), \text{ for } \lambda\in(\mathbb{Z}/N^2\mathbb{Z})^\times,$$

where P' is a point of order N, point Q' is a point of order N^2, and $e_N(P',[N]Q')=\zeta_N$. The condition on the Weil pairing is also satisfied on the mapped points because, using the properties of the Weil pairing, we have:

$$\begin{aligned} e_N([\lambda]P',[\lambda^{-1}]([N]Q')) &= e_N(P',[\lambda^{-1}]([N]Q'))^\lambda\\ &= e_N(P',[N]Q')^{\lambda\lambda^{-1}}\\ &= (\zeta_N)^{1+kN^2},\ k\in\mathbb{Z} \quad (\text{multiplication is in } (\mathbb{Z}/N^2\mathbb{Z})^\times)\\ &= \zeta_N = e_N(P',[N]Q'), \end{aligned}$$

so the automorphism λ is well-defined. Keeping the same notation, the restriction of the automorphism $\lambda\in(\mathbb{Z}/N^2\mathbb{Z})^\times$ on $X_1(N)$ is the map:

$$\lambda\colon (E,P)\mapsto (E,[\lambda]P), \quad \text{for} \quad \lambda\in(\mathbb{Z}/N\mathbb{Z})^\times,$$

where E and P are as before. Essentially, λ is reduced modulo N and then applied to the point of order N.

Let G be a group generated by all those automorphisms, i.e.

$$G = \langle \alpha, t, \omega, \lambda \in (\mathbb{Z}/N^2\mathbb{Z})^\times \rangle,$$

and let H be a subgroup of G generated by the same automorphisms except ω, i.e.

$$H = \langle \alpha, t, \lambda \in (\mathbb{Z}/N^2\mathbb{Z})^\times \rangle.$$

We are interested in the connection between the groups G and H and functions from $X_0(N)$. This connection is summarized in the following lemma.

Lemma 5.1.1. Let $N \geq 5$ be a positive integer. Let E be an elliptic curve over the field k such that $\mathrm{char}(k) \nmid N$. Let R be a point of order N^2 on the same curve. Let P and P' be points of order N and such that $P = [N]R$ and $e_N(P', [N]R) = \zeta_N$. Let G be a group defined by

$$G = \langle \alpha, t, \omega, \lambda \in (\mathbb{Z}/N^2\mathbb{Z})^\times \rangle,$$

and let H be a subgroup of G defined by

$$H = \langle \alpha, t, \lambda \in (\mathbb{Z}/N^2\mathbb{Z})^\times \rangle,$$

where

$$\begin{aligned}
\alpha &: (E, P', R) \mapsto (E, P' + [N]R, R), \\
t &: (E, P', R) \mapsto (E, P', [N+1]R), \\
\omega &: (E, P', R) \mapsto (E, P', P' + R), \\
\lambda &: (E, P', R) \mapsto (E, [\lambda]P', [\lambda^{-1}]R),\ \lambda \in (\mathbb{Z}/N^2\mathbb{Z})^\times.
\end{aligned}$$

Let $\Phi_N^*, \pi_{1,N}^*, \pi_{2,N}^*$ and ψ_N^* be pullback operators

$$\begin{aligned}
\Phi_N^* &: \overline{k}(X_1(N^2)) \to \overline{k}(X(N, N^2)), \text{ of the map } \Phi_N((E, P', R)) = (E, R), \\
\pi_{1,N}^* &: \overline{k}(X_1(N)) \to \overline{k}(X_1(N^2)), \text{ of the map } \pi_{1,N}((E, R)) = (E, [N]R), \\
\pi_{2,N}^* &: \overline{k}(X_1(N)) \to \overline{k}(X_1(N^2)), \text{ of the map } \pi_{2,N}((E, R)) = (E/\langle [N]R \rangle, R + \langle [N]R \rangle), \text{ and} \\
\psi_N^* &: \overline{k}(X_0(N)) \to \overline{k}(X_1(N)), \text{ of the map } \psi_N((E, P)) = (E, \langle P \rangle).
\end{aligned}$$

The function field $\Phi_N^*(\pi_{1,N}^*(\psi_N^*(\overline{k}(X_0(N)))))$, i.e. every pullback of a function from $\overline{k}(X_0(N)$ by the composition of those operators, is invariant under the group G. The function field $\Phi_N^*(\pi_{2,N}^*(\psi_N^*(\overline{k}(X_0(N)))))$ is invariant under the subgroup H.

Proof. Let (E,P',R) be an enhanced elliptic curve for $X(N,N^2)$. The pullback operators are mapping this point as:

$$(E,P',R) \overset{\Phi_N}{\longmapsto} (E,R) \overset{\pi_{1,N}}{\longmapsto} (E,[N]R) = (E,P) \overset{\psi_N}{\longmapsto} (E,\langle P\rangle) \quad \text{and}$$
$$(E,P',R) \overset{\Phi_N}{\longmapsto} (E,R) \overset{\pi_{2,N}}{\longmapsto} (E/\langle [N]R\rangle, R+\langle [N]R\rangle) = (E/\langle P\rangle, R+\langle P\rangle) \overset{\psi_N}{\longmapsto} (E/\langle P\rangle, \langle R+\langle P\rangle\rangle).$$

If the functions from $\bar{k}(X_0(N))$ are invariant under the group G, its generators should, combined with the pullbacks, produce the same enhanced elliptic curve $(E,\langle P\rangle)$ (or $(E/\langle P\rangle, R+\langle P\rangle)$ for the generators of the subgroup H) for $X_0(N)$.

We start with the generators of the group G. For α we have:

$$(E,P',R) \overset{\alpha}{\longmapsto} (E,P'+[N]R,R) \overset{\Phi_N}{\longmapsto} (E,R) \overset{\pi_{1,N}}{\longmapsto} (E,[N]R) = (E,P) \overset{\psi_N}{\longmapsto} (E,\langle P\rangle).$$

For t we have:

$$\begin{aligned}(E,P',R) &\overset{t}{\longmapsto} (E,P',[N+1]R) \overset{\Phi_N}{\longmapsto} (E,[N+1]R)\\ &\overset{\pi_{1,N}}{\longmapsto} (E,[N(N+1)]R) = (E,[N]R) = (E,P) \text{ (because } R \text{ is of order } N^2)\\ &\overset{\psi_N}{\longmapsto} (E,\langle P\rangle).\end{aligned}$$

For ω we have:

$$\begin{aligned}(E,P',R) &\overset{\omega}{\longmapsto} (E,P',P'+R) \overset{\Phi_N}{\longmapsto} (E,P'+R)\\ &\overset{\pi_{1,N}}{\longmapsto} (E,[N](P'+R)) = (E,[N]R) = (E,P) \text{ (because } P' \text{ is of order } N)\\ &\overset{\psi_N}{\longmapsto} (E,\langle P\rangle).\end{aligned}$$

And for $\lambda \in (\mathbb{Z}/N^2\mathbb{Z})^\times$ we have:

$$\begin{aligned}(E,P',R) &\overset{\lambda}{\longmapsto} (E,[\lambda]P',[\lambda^{-1}]R) \overset{\Phi_N}{\longmapsto} (E,[\lambda^{-1}]R)\\ &\overset{\pi_{1,N}}{\longmapsto} (E,[N]([\lambda^{-1}]R)) = (E,[\lambda^{-1}]P)\\ &\overset{\psi_N}{\longmapsto} (E,\langle [\lambda^{-1}]P\rangle) = (E,\langle P\rangle) \text{ (because } [\lambda^{-1}]P \in \langle P\rangle).\end{aligned}$$

Furthermore, for the generator α of the subgroup H we have:

$$\begin{aligned}(E,P',R) &\overset{\alpha}{\longmapsto} (E,P'+[N]R,R) \overset{\Phi_N}{\longmapsto} (E,R)\\ &\overset{\pi_{2,N}}{\longmapsto} (E/\langle [N]R\rangle, R+\langle [N]R\rangle) = (E/\langle P\rangle, R+\langle P\rangle)\\ &\overset{\psi_N}{\longmapsto} (E/\langle P\rangle, \langle R+\langle P\rangle\rangle).\end{aligned}$$

For the generator t we have:

$$\begin{aligned}(E,P',R) &\overset{t}{\mapsto} (E,P',[N+1]R) \overset{\Phi_N}{\longmapsto} (E,[N+1]R)\\ &\overset{\pi_{2,N}}{\longmapsto} (E/\langle[N(N+1)]R\rangle, R+\langle[N(N+1)]R\rangle) = (E/\langle P\rangle, R+\langle P\rangle)\\ &\overset{\psi_N}{\longmapsto} (E/\langle P\rangle, \langle R+\langle P\rangle\rangle).\end{aligned}$$

For the generator λ we have:

$$\begin{aligned}(E,P',R) &\overset{\lambda}{\mapsto} (E,[\lambda]P',[\lambda^{-1}]R) \overset{\Phi_N}{\longmapsto} (E,[\lambda^{-1}]R)\\ &\overset{\pi_{2,N}}{\longmapsto} (E/\langle[N]([\lambda^{-1}]R)\rangle, R+\langle[N]([\lambda^{-1}]R)\rangle) = (E/\langle[\lambda^{-1}]P\rangle, R+\langle[\lambda^{-1}]P\rangle)\\ &\overset{\psi_N}{\longmapsto} (E/\langle[\lambda^{-1}]P\rangle, \langle R+\langle[\lambda^{-1}]P\rangle\rangle) = (E/\langle P\rangle, \langle R+\langle P\rangle\rangle).\end{aligned}$$

■

Lemma 5.1.1 implies that the group G is isomorphic to the Galois group of the field extension $\overline{k}(X(N,N^2))/\overline{k}(X_0(N))$ and thus the equality $\overline{k}(X_0(N)) = \overline{k}(X(N,N^2))^G$ holds.

The subgroup H from Lemma 5.1.1 can be used to define a function field $k' := \overline{k}(X(N,N^2))^H$, which is an intermediate field such that

$$\overline{k}(X_0(N)) \subset k' \subset \overline{k}(X(N,N^2)),$$

and a function field for some modular curve. All the functions from the set $\Phi_N(\pi_{2,N}^*(\psi_N^*(k(X_0(N)))))$ are well-defined on that modular curve, due to their invariant property under the subgroup H.

Using the setup and the proof of Lemma 5.1.1, and the discussion above, we can prove the following theorem.

Theorem 5.1.2. Let $N \geq 5$ be an integer. Let H be a group generated by the automorphisms α, t and $\lambda \in (\mathbb{Z}/N^2\mathbb{Z})^\times$ i.e. a subgroup of $G = \langle \alpha, t, \omega, \lambda \in (\mathbb{Z}/N^2\mathbb{Z})^\times \rangle$. The subgroup H is not a normal subgroup of G.

Proof. The subgroup H will be a normal subgroup of G if for every $g \in G$ we have $gHg^{-1} = H$. To show that H is not a normal subgroup, it is enough to prove that the automorphism ω does not normalize the subgroup H, that is $\omega H \omega^{-1} \neq H$, i.e. that there exist $\lambda \in (\mathbb{Z}/N^2\mathbb{Z})^\times$, such that for an enhanced elliptic curve (E,P',Q') for $X(N,N^2)$,

the composition $\omega \circ \lambda \circ \omega^{-1}(E,P',Q')$ is not an element of the subgroup H. If $\omega \circ \lambda \circ \omega^{-1}(E,P',Q')$ is in H, then, because H is a subgroup, $\lambda^{-1} \circ \omega \circ \lambda \circ \omega^{-1}(E,P',Q')$ will also be in H. Therefore,

$$\begin{aligned}
\lambda^{-1} \circ \omega \circ \lambda \circ \omega^{-1}(E,P',Q') &= \lambda^{-1}(\omega(\lambda(E,P',Q'-P'))) \\
&= \lambda^{-1}(\omega(E,[\lambda]P',[\lambda^{-1}](Q'-P'))) \\
&= \lambda^{-1}(E,[\lambda]P',[\lambda^{-1}](Q'-P')+[\lambda]P') \\
&= (E,P',Q'-P'+[\lambda^2]P') = (E,Q'+[\lambda^2-1]P') \\
&= \omega^{\lambda^2-1}(E,P',Q').
\end{aligned}$$

From this equality, we can conclude that H will be a normal subgroup of the group G only if ω^{λ^2-1} acts as an identity for every λ. Take $\lambda = 2$. Then $\omega^{\lambda^2-1}(E,P',Q') = (E,P',3P'+Q') \neq (E,P',Q')$, because P' is of order $N > 3$. To conclude, H is not a normal subgroup of G. ■

Let the group G and pullbacks $\Phi_N^*, \pi_{1,N}^*, \pi_{2,N}^*, \psi_N^*$ be defined as in Lemma 5.1.1, and let k' be an intermediate field $\overline{k}(X_0(N)) \subset k' \subset \overline{k}(X(N,N^2))$ as discussed before. Working with the function fields shown in Figure 5.2, to get radical isogeny formulas on $X_0(N)$, we need to find an $\alpha \in \overline{k}(X_0(N))$ such that

$$\overline{k}(X_0(N))(\sqrt[N]{\alpha}) = \overline{k'}.$$

Functions from $\overline{k}(X_0(N))$ are identified with the composition of pullbacks $\Phi_N^*, \pi_{1,N}^*$ and ψ_N^*, which implies that α should be an element of the field $\Phi_N^*(\pi_{1,N}^*(\psi_N^*(\overline{k}(X_0(N)))))$. If such α exists, the field extension $\overline{k'}/\overline{k}(X_0(N))$ should be a cyclic extension, i.e. it should be a Galois extension, meaning that H should be a normal subgroup of G. In Theorem 5.1.2 we showed that this is not true, thus the extension $\overline{k'}/\overline{k}(X_0(N))$ cannot be a Galois extension. Similar to before, instead of the field $\overline{k}$ one could calculate the field of definition of automorphism with added automorphism λ included, and in that case work with a smaller field than $\overline{k}$.

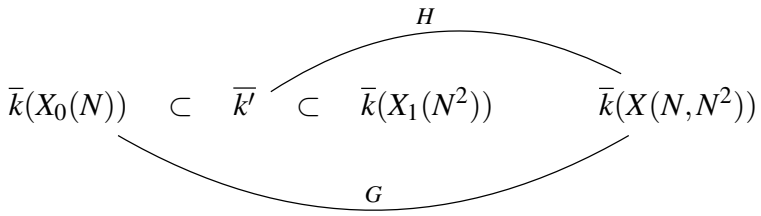

Figure 5.2: Function fields related to groups G and H

Returning to Example 3.2.5, the existence of radical isogeny formulas on $S_0(5)$ depends on finding a parametrization of $S_0(5)$ for which the extension $\mathbb{Q}(\zeta_5)(\beta')/\mathbb{Q}(\zeta_5)(\beta)$ is Galois. However, as we can see in the discussion above, such Galois extension is not possible in a more generalized setting of modular curves, and as a consequence, we have the following corollary.

Corollary 5.1.3. Let $N \geq 5$. Radical isogeny formulas on $S_0(N)$ are not possible.

CONCLUSION

This thesis focused on exploring the relationship between radical isogenies and modular curves. In this concluding section, we will give a brief overview of the main results, presented previously in Chapters 4 and 5.

In Chapter 4, we developed a generalized approach to radical isogeny formulas using the modular curve $X_1(N)$. To achieve this, we used enhanced elliptic curves for a congruence subgroup $\Gamma_1(N)$ for different N and maps between them. As previously explained in Chapter 3, radical isogenies are formulas used to calculate an isogeny, or a chain of isogenies of the same degree, starting from the elliptic curve E over a field k and a point on that curve. Given E in the Tate normal form with coefficients b and c, and a point P of order $N \geq 4$ on E, using radical isogenies we can calculate the coordinates of a new point P' such that the composition $E \xrightarrow{\varphi} E' \to E'/\langle P' \rangle$ is a cyclic isogeny of degree N^2. Formulas for the coordinates of the point P', as well as the coefficients b' and c' of the Tate normal form for elliptic curve $E'/\langle P' \rangle$, depend on b, c and some radicand $\rho \in k(b,c)$, that is a N-th root of a rational expression in the coefficients b and c. To generalize the concept of radical isogenies, our first step was to demonstrate that b, c, and ρ can be regarded as functions on the set of equivalence classes of enhanced elliptic curves. To complete the generalization process, in Section 4.1.1, we identified the smallest field of definition for these functions and established a field equality that implied the connection between coefficients b, c and b', c' in terms of enhanced elliptic and modular curves.

In Chapter 5, we pursued further generalization by considering the modular curve $X_0(N)$. This generalization could help find a better or more optimized radical isogeny formula, as explained in Example 3.2.5. We again utilized the maps between enhanced elliptic curves, now for a congruence subgroup $\Gamma_0(N)$, and explored the conditions under which radical isogeny formulas could exist. The central question in this context was

whether a particular field extension is Galois or not. Using the result from Theorem 5.1.2, we showed that the field in question cannot be Galois, and this implied that the existence of radical isogenies formulas on $X_0(N)$ for $N \geq 5$ is not possible.

BIBLIOGRAPHY

[1] Ajtai, M.: *Generating hard instances of lattice problems (extended abstract)*. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing - STOC '96*, STOC '96, pages 99–108. ACM Press, 1996. http://dx.doi.org/10.1145/237814.237838. ↑ 44.

[2] Alagic, Gorjan, Jacob Alperin-Sheriff, Daniel Apon, David Cooper, Quynh Dang, Yi Kai Liu, Carl Miller, Dustin Moody, Rene Peralta, Ray Perlner, Angela Robinson, and Daniel Smith-Tone: *Status report on the first round of the NIST post-quantum cryptography standardization process*. National Institute of Standards and Technology, 2019. http://dx.doi.org/10.6028/NIST.IR.8240. ↑ 43.

[3] Alagic, Gorjan, Daniel Apon, David Cooper, Quynh Dang, Thinh Dang, John Kelsey, Jacob Lichtinger, Yi Kai Liu, Carl Miller, Dustin Moody, Rene Peralta, Ray Perlner, Angela Robinson, and Daniel Smith-Tone: *Status report on the third round of the NIST Post-Quantum Cryptography Standardization process*. National Institute of Standards and Technology (U.S.), 2022. http://dx.doi.org/10.6028/NIST.IR.8413. ↑ 44.

[4] Albrecht, Martin R., Daniel J. Bernstein, Tung Chou, Carlos Cid, Jan Gilcher, *et al.*: *Classic McEliece: conservative code-based cryptography*, 2022. https://csrc.nist.gov/Projects/post-quantum-cryptography/round-4-submissions, Accessed: 2024-02-17. ↑ 45.

[5] Aragon, Nicolas, Paulo Barreto, Slim Bettaieb, Loic Bidoux, Olivier Blazy, *et al.*: *BIKE: bit flipping key encapsulation*, 2022. https://csrc.nist.gov/Projects/

post-quantum-cryptography/round-4-submissions, Accessed: 2024-02-17. ↑ 45.

[6] Aumasson, Jean Philippe, Daniel J Bernstein, Ward Beullens, Christoph Dobraunig, Maria Eichlseder, *et al.*: *SPHINCS+: Submission to the NIST post-quantum project, v.3*, 2020. https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022, Accessed: 2024-02-17. ↑ 45.

[7] Avanzi, Roberto, Joppe Bos, Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, John M. Schanck, Peter Schwabe, Gregor Seiler, and Damien Stehlé: *CRYSTALS-KYBER algorithm specifications and supporting documentation (version 3.0)*, 2020. https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022, Accessed: 2024-02-17. ↑ 44.

[8] Azarderakhsh, Reza, Matthew Campagna, Craig Costello, Luca De Feo, Basil Hess, *et al.*: *Supersingular isogeny key encapsulation*, 2022. https://csrc.nist.gov/Projects/post-quantum-cryptography/round-4-submissions, Accessed: 2024-02-17. ↑ 45.

[9] Bai, Shi, Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, Peter Schwabe, Gregor Seiler, and Damien Stehlé: *CRYSTALS-DILITHIUM: Algorithm specifications and supporting documentation*, 2020. https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022, Accessed: 2024-02-17. ↑ 44.

[10] Benioff, Paul: *The computer as a physical system: A microscopic quantum mechanical hamiltonian model of computers as represented by turing machines*. Journal of Statistical Physics, 22(5):563–591, 1980, ISSN 1572-9613. http://dx.doi.org/10.1007/BF01011339. ↑ 42.

[11] Bernstein, Daniel J, Johannes Buchmann, and Erik Dahmen (editors): *Post-Quantum Cryptography*. Springer Berlin Heidelberg, 2009, ISBN 9783540887027. http://dx.doi.org/10.1007/978-3-540-88702-7. ↑ 44, 45.

[12] Bernstein, Daniel J., Luca De Feo, Antonin Leroux, and Benjamin Smith: *Faster computation of isogenies of large prime degree*. Open Book Series, 4(1):39–55, 2020, ISSN 2329-9061. http://dx.doi.org/10.2140/obs.2020.4.39. ↑ 2.

[13] Beullens, Ward: *Breaking Rainbow takes a weekend on a laptop*. In *Lecture Notes in Computer Science*, page 464–479. Springer Nature Switzerland, 2022, ISBN 9783031159794. http://dx.doi.org/10.1007/978-3-031-15979-4_16. ↑ 45.

[14] Box, Josha: *Computing models for quotients of modular curves*. Research in Number Theory, 7(3), 2021, ISSN 2363-9555. http://dx.doi.org/10.1007/s40993-021-00276-8. ↑ 33.

[15] Castryck, Wouter and Thomas Decru: *An efficient key recovery attack on SIDH*. In *Lecture Notes in Computer Science*, page 423–447. Springer Nature Switzerland, 2023, ISBN 9783031305894. http://dx.doi.org/10.1007/978-3-031-30589-4_15. ↑ 1, 45, 53.

[16] Castryck, Wouter, Thomas Decru, Marc Houben, and Frederik Vercauteren: *Horizontal racewalking using radical isogenies*. In *Lecture Notes in Computer Science*, page 67–96. Springer Nature Switzerland, 2022, ISBN 9783031229664. http://dx.doi.org/10.1007/978-3-031-22966-4_3. ↑ 2, 57.

[17] Castryck, Wouter, Thomas Decru, and Frederik Vercauteren: *Radical isogenies*. In *Lecture Notes in Computer Science*, page 493–519. Springer International Publishing, 2020, ISBN 9783030648343. http://dx.doi.org/10.1007/978-3-030-64834-3_17. ↑ ii, iii, 2, 54, 55, 56, 57, 74.

[18] Castryck, Wouter, Tanja Lange, Chloe Martindale, Lorenz Panny, and Joost Renes: *CSIDH: An efficient post-quantum commutative group action*. In *Advances in Cryptology – ASIACRYPT 2018*, page 395–427. Springer International Publishing, 2018, ISBN 9783030033323. http://dx.doi.org/10.1007/978-3-030-03332-3_15. ↑ 1, 48, 51, 53.

[19] Cervantes-Vázquez, Daniel, Mathilde Chenu, Jesús Javier Chi-Domínguez, Luca De Feo, Francisco Rodríguez-Henríquez, and Benjamin Smith: *Stronger and faster side-channel protections for CSIDH*. In *Lecture Notes in Computer Science*, page 173–193. Springer International Publishing, 2019, ISBN 9783030305307. http://dx.doi.org/10.1007/978-3-030-30530-7_9. ↑ 2.

[20] Charles, Denis X., Kristin E. Lauter, and Eyal Z. Goren: *Cryptographic hash functions from expander graphs*. Journal of Cryptology, 22(1):93–113, September 2007, ISSN 1432-1378. http://dx.doi.org/10.1007/s00145-007-9002-x. ↑ 2, 48.

[21] Chen, Cong, Jeffrey Hoffstein, Andreas Hülsing, Joost Rijneveld, John M Schanck, *et al.*: *NTRU: algorithm specifcations and supporting documentation*, 2020. https://csrc.nist.gov/Projects/post-quantum-cryptography/post-quantum-cryptography-standardization/round-3-submissions, Accessed: 2024-02-17. ↑ 44.

[22] Cornell, Gary, Joseph H. Silverman, and Glenn Stevens (editors): *Modular Forms and Fermat's Last Theorem*. Springer New York, 1997, ISBN 9781461219743. http://dx.doi.org/10.1007/978-1-4612-1974-3.

[23] Costello, Craig and Huseyin Hisil: *A simple and compact algorithm for SIDH with arbitrary degree isogenies*. In *Lecture Notes in Computer Science*, page 303–329. Springer International Publishing, 2017, ISBN 9783319706979. http://dx.doi.org/10.1007/978-3-319-70697-9_11. ↑ 2.

[24] Costello, Craig and Benjamin Smith: *Montgomery curves and their arithmetic: The case of large characteristic fields*. Journal of Cryptographic Engineering, 8(3):227–240, 2017, ISSN 2190-8516. http://dx.doi.org/10.1007/s13389-017-0157-6. ↑ 58, 59.

[25] Couveignes, Jean Marc: *Hard homogeneous spaces*. Cryptology ePrint Archive, Paper 2006/291, 2006. https://eprint.iacr.org/2006/291. ↑ 1.

[26] Cox, David A: *Primes of the Form $x^2 + ny^2$: Fermat, Class Field Theory, and Complex Multiplication. with Solutions*, volume 387. American Mathematical Soc., 2022. ↑ 26.

[27] De Feo, Luca: *Mathematics of isogeny based cryptography*, 2017. https://arxiv.org/abs/1711.04062. ↑ 46, 48.

[28] De Feo, Luca, David Kohel, Antonin Leroux, Christophe Petit, and Benjamin Wesolowski: *SQISign: compact post-quantum signatures from quaternions and isogenies*. In *Lecture Notes in Computer Science*, page 64–93. Springer International Publishing, 2020, ISBN 9783030648374. http://dx.doi.org/10.1007/978-3-030-64837-4_3. ↑ 1.

[29] De Feo, Luca, Antonin Leroux, Patrick Longa, and Benjamin Wesolowski: *New algorithms for the Deuring correspondence: Towards practical and secure SQISign signatures*. In *Lecture Notes in Computer Science*, page 659–690. Springer Nature Switzerland, 2023, ISBN 9783031305894. http://dx.doi.org/10.1007/978-3-031-30589-4_23. ↑ 1.

[30] Deligne, P. and M. Rapoport: *Les schémas de modules de courbes elliptiques*. In *Modular Functions of One Variable II*, page 143–316. Springer Berlin Heidelberg, 1973, ISBN 9783540378556. http://dx.doi.org/10.1007/978-3-540-37855-6_4. ↑ 69.

[31] Diamond, Fred and Jerry Shurman: *A first course in modular forms*. In *Graduate Texts in Mathematics*. Springer, 2005, ISBN 9780387272269. http://dx.doi.org/10.1007/978-0-387-27226-9. ↑ 5, 31, 67, 69.

[32] Diffie, W. and M. Hellman: *New directions in cryptography*. IEEE Transactions on Information Theory, 22(6):644–654, 1976, ISSN 0018-9448. http://dx.doi.org/10.1109/TIT.1976.1055638. ↑ 36, 37.

[33] Ding, Jintai, Chen Ming-Shing, Matthias Kannwischer, Jacques Patarin, Albrecht Petzoldt, *et al.*: *Rainbow: algorithm specification and documentation*, 2020. https://csrc.nist.gov/Projects/post-quantum-cryptography/post-

quantum-cryptography-standardization/round-3-submissions, Accessed: 2024-02-17. ↑ 45.

[34] Dose, Valerio, Guido Lido, and Pietro Mercuri: *Automorphisms of Cartan modular curves of prime and composite level*. Algebra & Number Theory, 16(6):1423–1461, 2022, ISSN 1937-0652. http://dx.doi.org/10.2140/ant.2022.16.1423. ↑ 33, 76, 80, 81.

[35] Dujella, Andrej: *Teorija brojeva*. Školska knjiga, Zagreb, 2019. ↑ 15.

[36] Dworkin, Morris J: *Advanced Encryption Standard (AES)*. National Institute of Standards and Technology (U.S.), 2023. http://dx.doi.org/10.6028/nist.fips.197-upd1. ↑ 35.

[37] Fouque, Pierre Alain, Jeffrey Hoffstein, Paul Kirchner, Vadim Lyubashevsky, Thomas Pornin, Thomas Prest, Thomas Ricosset, Gregor Seiler, William Whyte, and Zhenfei Zhang: *Falcon: fast-Fourier lattice-based compact signatures over NTRU, specification v1. 2*, 2020. https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022, Accessed: 2024-02-17. ↑ 44.

[38] Galbraith, Steven D.: *Mathematics of Public Key Cryptography*. Cambridge University Press, 2012, ISBN 9781107013926. http://dx.doi.org/10.1017/CBO9781139012843. ↑ 58, 59.

[39] Grover, Lov K.: *A fast quantum mechanical algorithm for database search*. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing - STOC '96*, STOC '96, pages 212–219. ACM Press, 1996. http://dx.doi.org/10.1145/237814.237866. ↑ 43.

[40] Hartshorne, Robin: *Algebraic Geometry*. Springer New York, 1977, ISBN 9781475738490. http://dx.doi.org/10.1007/978-1-4757-3849-0. ↑ 33.

[41] Jao, David and Luca De Feo: *Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies*. In *Lecture Notes in Computer Science*, page 19–34.

Springer Berlin Heidelberg, 2011, ISBN 9783642254055. http://dx.doi.org/10.1007/978-3-642-25405-5_2. ↑ 1.

[42] Katz, Nicholas M. and Barry Mazur: *Arithmetic Moduli of Elliptic Curves. (AM-108)*. Princeton University Press, 1985, ISBN 9781400881710. http://dx.doi.org/10.1515/9781400881710. ↑ 33, 69.

[43] Kim, Suhri, Kisoon Yoon, Young Ho Park, and Seokhie Hong: *Optimized method for computing odd-degree isogenies on Edwards curves*. In *Advances in Cryptology – ASIACRYPT 2019*, page 273–292. Springer International Publishing, 2019, ISBN 9783030346218. http://dx.doi.org/10.1007/978-3-030-34621-8_10. ↑ 2.

[44] Kipnis, Aviad, Jacques Patarin, and Louis Goubin: *Unbalanced Oil and Vinegar signature schemes*. In *Lecture Notes in Computer Science*, page 206–222. Springer Berlin Heidelberg, 1999, ISBN 9783540489108. http://dx.doi.org/10.1007/3-540-48910-X_15. ↑ 45.

[45] Koblitz, Neal: *Elliptic curve cryptosystems*. Mathematics of Computation, 48(177):203–209, 1987, ISSN 1088-6842. http://dx.doi.org/10.1090/S0025-5718-1987-0866109-5. ↑ 39.

[46] Lang, Serge: *Algebra*. Springer New York, 2002, ISBN 9781461300410. http://dx.doi.org/10.1007/978-1-4613-0041-0. ↑ 5, 9, 21.

[47] Maino, Luciano and Chloe Martindale: *An attack on SIDH with arbitrary starting curve*. Cryptology ePrint Archive, Paper 2022/1026, 2022. https://eprint.iacr.org/2022/1026. ↑ 1, 45, 53.

[48] McEliece, Robert J: *A public-key cryptosystem based on algebraic coding theory*. Technical report, NASA, 1978. https://ipnpr.jpl.nasa.gov/progress_report2/42-44/44N.PDF. ↑ 45.

[49] Melchor, Carlos Aguilar, Nicolas Aragon, Slim Bettaieb, Loıc Bidoux, Olivier Blazy, *et al.*: *Hamming quasi-cyclic (HQC)*, 2022. https://csrc.nist.gov/

Projects/post-quantum-cryptography/round-4-submissions, Accessed: 2024-02-17. ↑ 45.

[50] Merkle, Ralph C.: *A certified digital signature*. In *Lecture Notes in Computer Science*, page 218–238. Springer New York, 1989, ISBN 9780387973173. http://dx.doi.org/10.1007/0-387-34805-0_21. ↑ 45.

[51] Miller, Victor S.: *Use of elliptic curves in cryptography*. In *Advances in Cryptology — CRYPTO '85 Proceedings*, page 417–426. Springer Berlin Heidelberg, 1985, ISBN 9783540164630. http://dx.doi.org/10.1007/3-540-39799-X_31. ↑ 39, 59.

[52] Montgomery, Peter L: *Speeding the Pollard and elliptic curve methods of factorization*. Mathematics of computation, 48(177):243–264, 1987. ↑ 58.

[53] Moody, Dustin, Gorjan Alagic, Daniel C Apon, David A Cooper, Quynh H Dang, John M Kelsey, Yi Kai Liu, Carl A Miller, Rene C Peralta, Ray A Perlner, Angela Y Robinson, Daniel C Smith-Tone, and Jacob Alperin-Sheriff: *Status report on the second round of the NIST post-quantum cryptography standardization process*. National Institute of Standards and Technology, 2020. http://dx.doi.org/10.6028/NIST.IR.8309. ↑ 44.

[54] Onuki, Hiroshi and Tomoki Moriya: *Radical isogenies on Montgomery curves*. In *Lecture Notes in Computer Science*, page 473–497. Springer International Publishing, 2022, ISBN 9783030971212. http://dx.doi.org/10.1007/978-3-030-97121-2_17. ↑ ii, iii, 3, 4, 54, 58, 59, 60, 61.

[55] Onuki, Hiroshi and Tsuyoshi Takagi: *On collisions related to an ideal class of order 3 in CSIDH*. In *Lecture Notes in Computer Science*, page 131–148. Springer International Publishing, 2020, ISBN 9783030582081. http://dx.doi.org/10.1007/978-3-030-58208-1_8. ↑ 51.

[56] Panny, Lorenz: *Cryptography on isogeny graphs*. Eindhoven: Technische Universiteit Eindhoven, 2021. ↑ 47.

[57] Perez Broon, Fouazou Lontouo, Thinh Dang, Emmanuel Fouotsa, and Dustin Moody: *Isogenies on twisted Hessian curves*. Journal of Mathematical Cryptology, 15(1):345–358, 2021, ISSN 1862-2984. http://dx.doi.org/10.1515/jmc-2020-0037. ↑ 2.

[58] Pribanić, Valentina: *Radical isogenies and modular curves*. Advances in Mathematics of Communications, 2023, ISSN 1930-5338. http://dx.doi.org/10.3934/amc.2023019. ↑ ii, iii.

[59] Rivest, R. L., A. Shamir, and L. Adleman: *A method for obtaining digital signatures and public-key cryptosystems*. Communications of the ACM, 21(2):120–126, 1978, ISSN 1557-7317. http://dx.doi.org/10.1145/359340.359342. ↑ 39.

[60] Robert, Damien: *Breaking SIDH in polynomial time*, page 472–503. Springer Nature Switzerland, 2023, ISBN 9783031305894. http://dx.doi.org/10.1007/978-3-031-30589-4_17. ↑ 1, 45, 53.

[61] Rostovtsev, Alexander and Anton Stolbunov: *Public-key cryptosystem based on isogenies*. Cryptology ePrint Archive, Paper 2006/145, 2006. https://eprint.iacr.org/2006/145. ↑ 1.

[62] Shor, P.W.: *Algorithms for quantum computation: discrete logarithms and factoring*. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, SFCS-94. IEEE Comput. Soc. Press, 1994. http://dx.doi.org/10.1109/SFCS.1994.365700. ↑ 43.

[63] Siksek, Samir: *Explicit arithmetic of modular curves*, 2019. https://homepages.warwick.ac.uk/staff/S.Siksek/teaching/modcurves/lecturenotes.pdf, Accessed: 2023-12-29. ↑ 31.

[64] Silverman, Joseph H.: *The Arithmetic of Elliptic Curves*. Springer New York, 2009, ISBN 9780387094946. http://dx.doi.org/10.1007/978-0-387-09494-6. ↑ 5, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 30, 50, 51.

[65] Silverman, Joseph H, Jill Pipher, and Jeffrey Hoffstein: *An introduction to mathematical cryptography*. Springer New York, 2008, ISBN 9780387779942. http://dx.doi.org/10.1007/978-0-387-77993-5. ↑ 13, 35, 38, 39.

[66] Streng, Marco: *Generators of the group of modular units for* $\Gamma^1(N)$ *over the rationals*. Annales Henri Lebesgue, 6:95–116, 2023. http://dx.doi.org/10.5802/ahl.160. ↑ 15, 17.

[67] Vélu, Jacques: *Isogénies entre courbes elliptiques*. CR Acad. Sci. Paris, Séries A, 273:305–347, 1971. ↑ 2, 20.

[68] Voight, John: *Quaternion Algebras*. Springer International Publishing, 2021, ISBN 9783030566944. http://dx.doi.org/10.1007/978-3-030-56694-4. ↑ 50.

[69] Washington, Lawrence C: *Elliptic curves: number theory and cryptography*. CRC press, 2008. ↑ 5, 20, 25, 26.

[70] Wong, David: *Real-world cryptography*. Manning Publications Co., 2021, ISBN 9781617296710. ↑ 35, 42.

[71] Wroński, Michał: *Application of Velusqrt algorithm to Huff's curves*. Publicationes Mathematicae Debrecen, 100(Supplementum):639–653, 2022, ISSN 0033-3883. http://dx.doi.org/10.5486/PMD.2022.Suppl.6. ↑ 2.

CURRICULUM VITAE

Valentina Pribanić was born on September 15, 1990, in Ogulin, Croatia. She attended primary school in Tounj and pursued her secondary education in Ogulin. In 2009, she enrolled in the Undergraduate University Programme in Mathematics at the Mathematics Department of the Faculty of Science, University of Zagreb. Completing her undergraduate degree in 2012, she continued her studies in the Graduate Programme in Applied Mathematics at the same department, achieving her graduation *cum laude* in 2015. Her Master's thesis titled *NTRU Cryptosystem* was completed under the guidance of Professor Andrej Dujella.

In 2017, she enrolled in the Postgraduate Doctoral Study Programme in Mathematics, supervised by Professor Matija Kazalicki. She is a member of the *Seminar on Number Theory and Algebra*. Currently, she is employed as a department head in the Croatian Ministry of Interior.

She is the author of one paper accepted for publication:

- Pribanić, Valentina: *Radical isogenies and modular curves*, Advances in Mathematics of Communications, 2023, ISSN 1930-5338.

IZJAVA O IZVORNOSTI RADA

Ja, _____, student/ica Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu, s prebivalištem na adresi _____,OIB _____, JMBAG _____, ovim putem izjavljujem pod materijalnom i kaznenom odgovornošću da je moj završni/diplomski/doktorski rad pod naslovom:

_____

_____

_____, isključivo moje autorsko djelo, koje je u potpunosti samostalno napisano uz naznaku izvora drugih autora i dokumenata korištenih u radu.

U Zagrebu, _____

_____

Potpis