Conics and Elliptic Curves in Number Theory and Cryptography

Abdullatif Abdulrahman M Altheyab
Supervised by Dr. Alex Ghitza

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

The University of Melbourne
School of Mathematics and Statistics

17th October 2018

Abstract

The interplay between number theory and algebraic geometry has been fruitful for both fields. In this thesis, we study the groups arising from elliptic curves and pointed conics from both arithmetical and algebro-geometric perspectives, with an emphasis on finiteness results and the existence of different normal forms, as well as the broader conceptual foundations that underlie the subject matter. In general, notions and results that have applications to cryptography are given special attention, and the applications of the aforementioned groups to public-key cryptography constitutes the main topic of the final chapter.

Acknowledgements

All praise and glory to Allah who provided me the blessings and strength to complete this work.

I am very thankful for my supervisor, Dr. Alex Ghitza, whom was consistently able to help me with material that was at the right level and with guidance in regards to the way forward. I am deeply grateful for the time he has invested in my project, and for all the encouragement over the course of the last three semesters.

I would also like thank Melbourne University for nurturing me over these past two years and helping me rapidly expand my mathematical knowledge and ability and confidence with the English language. Thank you also the School of Mathematics and Statistics which has provided the facilities that have allowed me to study diligently and effectively.

A special thank you goes to my sponsor whose help has allowed me to travel here and study under some of the best mathematicians in Australia, and to the Saudi Arabian Cultural Mission for helping out with many of the difficulties associated with studying in a foreign country.

Much thanks, with love, to my wife Fatimah who has stood by me through some difficult times and never made me feel unwanted.

Thank you to my son Abdulrahman, whose endless energy has picked me up when times were hard. I promise I'll play with you more in the months after this is all over.

Thank you to my parents, and to my brothers and sisters, who encouraged me to come here, believed in me right from the beginning, and never stopped supporting me.

Thank you to my friends, whom I've had the fortune of learning advanced mathematics with over the past two years. The process would have been much less interesting and rewarding without you guys around to discuss our shared interests, and my English has improved immensely through our many discussions.

Contents

1 Fields, separability, and degree **9**

1.1 Set-theoretic ideals 9

1.2 Separable polynomials 11

1.3 Separable elements 13

1.4 Degree of a morphism of fields 16

1.5 Transcendence degree 17

2 Closure operators and Galois connections **19**

2.1 Closure Operators 19

2.2 Galois connections 23

3 Curves and varieties **29**

3.1 Affine space and the Zariski topology 29

3.2 Noetherian spaces and Irreducibility 33

3.3 Affine varieties and regular mappings 35

3.3.1 Localization and multiplicity 38

3.4 Dimension of an affine variety 40

3.5 Smoothness 41

3.6 Affine curves 42

3.7 Projective space, projective varieties and projective curves 42

3.8 Regular functions between projective curves 43

3.9 Regular differential forms and genus 45

3.10 Divisors and Picard groups 46

4 Elliptic Curves **49**

4.1 Isogenies 51

4.2 Planar elliptic curves 53

4.3 Weierstrass elliptic curves 56

4.4 Other normal forms for Elliptic curves 59
4.5 Separable and inseparable degree of isogenies 61
4.6 Ordinary versus supersingular elliptic curves 64
4.7 Velu's Formula 66
4.8 The group of $\mathbb{Q}$-rational points of an elliptic curve 69
4.9 The cardinality of $E(\mathbb{F}_q)$ 70

5 Pointed Conics **71**
5.1 Conics and the large matrix 71
5.2 The group law on a pointed conic 74
5.3 Arrows between pointed conics 76
5.4 Pell conics 80
5.5 The rational unit circle 83

6 Applications to Cryptography **85**
6.1 The Diffie-Hellman Key Exchange 86
6.2 Why does p have to be non-smooth? 88
6.3 The Diffie-Hellman key-exchange protocol over elliptic curves 90
6.4 The need for post-quantum cryptosystems 91
6.5 Supersingular isogeny-based Diffie-Helman (SIDH) key exchange 92

Appendices **95**

A Python code for Velu's Formula **96**

Introduction

The interactions between number theory and geometry have a long and distinguished history. Euclid's formula for Pythagorean triples, for example, has at it's heart a geometric insight about circle and lines, now referred to as stereographic projection, and in light of the ancient Greek preoccupation with conic sections, it's perhaps not surprising that a Greek name is associated with this formula.

The fruitful interplay between arithmetic and geometry continues to this day. Many of these connections center around *elliptic curves*, which are smooth projective curves of genus 1 equipped with a distinguished basepoint. Note that elliptic curves have only a passing connection to the ellipses of ancient Greek geometry; for example, the equation $y^2 = x^3 + 1$ defines an elliptic curve, but not an ellipse. Every elliptic curves becomes a group in a canonical way, and the study of the group structure has provided remarkable insights into number theoretic phenomena. A recent success story is Andrew Wile's proof of the modularity theorem for semistable elliptic curves, which was enough to imply Fermat's last theorem. Elliptic curves are the topic of Chapter 4.

Recently, Shirali [41] has shown that conics, too, become groups in a natural way, once a distinguished basepoint has been chosen. Chapter 5 examines these groups and studies different notions of morphism between them, emphasizing connections to number theory and arithmetic. Pell conics, which are defined by the equation $x^2 - dy^2 = 1$, receive particular attention, due to the availability of an explicit group law in this context.

Chapter 6 looks at some applications to public-key cryptography, and focuses in particular on *elliptic curve cryptography* (ECC). Since the invention of ECC in the mid 1980s, it has grown in popularity due to the comparatively shorter key lengths, and today enjoys widespread adoption [19]. Our emphasis is on key-exchange protocols, which allow parties that have never communicated before to establish a shared secret even if an eavesdropper can hear every message they exchange. The thesis concludes with a look at how elliptic curves are being used to develop key-exchange protocols that can resist attack by a quantum computer.

Chapters 1,2 and 3 are preliminary material and can probably be skipped by the impatient reader. Chapter 1 develops some aspects of field theory that are particularly relevant to the study of arithmetical geometry. Chapter 2 builds the basis theory of Galois connections and closure operators due to their wide applicability in connection with the Zariski topology and field theory. This leads, in some ways, to a more satisfying theory; for example, the fact that the Zariski-closure of a subset X of affine space (defined as the intersection of all $Y \supseteq X$ such that Y is Zariski-closed) exactly equals $V(I(X))$, is an immediate consequence of general facts about Galois connections that we might as well prove in maximum generality. Chapter 3 establishes the basic facts about affine and projective varieties that are used in the rest of the thesis.

Chapters 3 and 4 roughly follow Joseph H. Silverman's *The Arithmetic of Elliptic Curves* [43]. Silverman's book is very interesting insofar as it goes to great lengths to avoid technical machinery and abstract generalities; for example, his definition of the Picard group of a curve avoids mention of line bundles altogether in favour of the more elementary definition in terms of equivalence classes of formal $\mathbb{Z}$-linear combinations of points. This aversion to technical machinery is greatly reflected in the present work, in which, for example, no mention of ringed spaces is made. Hopefully, this makes the material more accessible to a wider audience.

Notation and Terminology

The natural numbers

The set of natural numbers $\mathbb{N}$ will begin at zero:

$$\mathbb{N} = \{0, 1, 2, \ldots\}.$$

The main relevance of this for our purposes is that it makes the following statements true.

(a) A vector space is finite-dimensional iff it has a basis whose cardinality is a natural number.

(b) The kernel of a group homomorphism is finite iff the cardinality of its kernel is a natural number.

(c) The Krull dimension of a variety is always a natural number.

(d) If P is a regular function, then the value of $\operatorname{ord}_P(f)$ is either a natural number or infinity.

Cardinality

If S is a set, we'll write $\#S$ for the cardinality of S. This notation will usually be employed where $\#S$ is a natural number, however in full generality it should be regarded as a cardinal number. The advantage of using cardinal numbers is that they allow us to phrase Proposition 4.5.2 and Proposition 4.5.4 without adding finiteness assumptions, which is arguably more satisfying.

Similarly, the dimension of a vector space (i.e. a K-module) will, in general, be regarded as a cardinal number. This allows us to speak of the dimension of the field $K(x)$ as a K-module, for example; its dimension is $\aleph_0$, the smallest infinite cardinal number.

Fields

If K is a field, write $\overline{K}$ for its algebraic closure and K^* for its group of units. Given $a \in K$, write $\overline{a}$ for the image of a under the inclusion $K \hookrightarrow \overline{K}$. Similarly, if $f \in K[x]$ is a polynomial, the corresponding element of $\overline{K}[x]$ will be denoted $\overline{f}$.

Rings versus algebras

For the purposes of this report, every K*-algebra* will be associative and unital by default, but not necessarily commutative, and a K*-ring* will be a commutative k-algebra. We'll use the term *algebra* to mean a $\mathbb{Z}$-algebra (i.e. a not-necessarily commutative ring) and reserve the term *ring* to mean a $\mathbb{Z}$-ring (i.e. all our rings will be commutative.) The main advantage of these conventions is that they allow us to avoid the phrase 'non-commutative ring,' since these are just called algebras.

The above changes also fit better with the definition of a ringed space, since the rings of the structure sheaf are always assumed commutative, and although the notion of a ringed space is not used in this thesis, it is good to prepare for the future in small ways such as this. Note also that the above conventions are consistent with the standard meaning of phrases like 'a K-scheme is a locally K-ringed space X such that...' in which the term 'K-ringed' implicitly carries an assumption of commutativity, which is consistent with the use of the term K-ring as introduced above.

Projections and indeterminates

The notation π_n will be overloaded in the following way. Firstly, it will denote the projection to the ith factor in a Cartesian product. For example:

$$\pi_1 : A \times B \to A, \qquad \pi_2 : A \times B \to B.$$

The notation π_i will also be used to denote indeterminates in a polynomial ring whenever a canonical ordering on the variables is desired. In particular, if $f \in K[\pi_1, \pi_2]$, then we can write $f(x, y)$ without ambiguity; it is understood to mean the result of replacing every copy of π_1 with x and every copy of π_2 with y. For instance, if $f \in K[\pi_1, \pi_2]$ is given by $f = \pi_1^2 + \pi_2^2$, then $f(x, y) = x^2 + y^2$ and $f(3, 4) = 25$.

Affine space

Given a natural number n, we write $\mathbb{A}^n_K$ for the set $\overline{K}^n$. This means in particular that, as sets $\mathbb{A}^n_K$ and $\mathbb{A}^n_{\overline{K}}$ mean the same thing. However, as topological spaces, they're different, an in particular, the latter has more closed sets than the former. See Definition 3.1.4 for the precise definition.

Chapter 1

Fields, separability, and degree

1.1 Set-theoretic ideals

In ring theory, ideals are characterized by two conditions; they're closed under taking multiples, and also under addition. Subsets satisfying only the first of these conditions are also useful. For example, the set of all elements of $K[x]$ that have a degree-1 factor is closed under taking multiples, but not closed under addition; consider the sum of x and $1 - x$, for example. The following definition provides terminology for this more general notion.

Definition 1.1.1. Let R denote a ring (or even just a commutative monoid). Then a *set-theoretic ideal* of R is a set $S \subseteq R$ that is closed under multiples, meaning that if $x \in S$, then for all $a \in R$ we have $ax \in S$.

Remark. The above notion is sometimes referred to as a 'monoid ideal,' but the phrase set-theoretic ideal is better. In particular, note that rings are monoid objects in the world of $\mathbb{Z}$-modules, whereas monoids are monoid objects in the world of sets. In light of this, the thing that's really changing is not whether we're dealing with monoids, it's whether we're working in **Set** or **Ab**. It is consequently better to choose terminology that emphasizes this.

There are many natural examples of set-theoretic ideals that are not closed under addition. We've already discussed that the notion of having a degree-1 factor. Another, more basic example is the set of all zero divisors in a ring (if x is a zero-divisor, then so too is ax.) Note that in the ring $\mathbb{Z}/6$, both 2 and 3 are zero divisors, but their sum, being its own multiplicative inverse, is not. Another example of a set-theoretic ideal is the set of all non-units in a ring (if x is a non-unit in a ring, then so too is ax.) Once again, the sum of two non-units needn't be a non-unit; for example, consider 3 and -2 in $\mathbb{Z}$.

Proposition 1.1.2. Let R denote a ring (or even just a commutative monoid) and suppose $S \subseteq R$ is a subset. Then the following are equivalent:

(a) S is a set-theoretic ideal

(b) $RS \subseteq S$

(c) $RS = S$

Proof. For (a) → (b), suppose S is a set-theoretic ideal. Consider $x \in RS$. We must show that $x \in S$. Write $x = ay$ for some $a \in R$ and some $y \in S$. Since $y \in S$, thus $ay \in S$. Hence $x \in S$, as desired. For (b) → (a), suppose $RS \subseteq S$. Consider $x \in S$. We must show that for all $a \in R$, we have $ax \in S$. Since $a \in R$ and $x \in S$, thus $ax \in RS$. So $ax \in S$, as desired. For (b) → (c), suppose $RS \subseteq S$. We must show $RS \supseteq S$. Since $R \supseteq \{1\}$, hence $RS \supseteq 1S$. Hence $RS \supseteq S$, as desired. The implication (c) → (b) is trivial. □

Proposition 1.1.3. Given a ring R and a subset $S \subseteq R$, the set

$$RS := \{rx : r \in R, x \in S\}$$

is the smallest set-theoretic ideal containing S.

Proof. We must show the following:

(a) RS is a set-theoretic ideal

(b) $RS \supseteq S$

(c) If $X \supseteq S$ is a set-theoretic ideal, then $X \supseteq RS$.

For part (a), we must show that $R(RS) \subseteq RS$. But since $RR \subseteq R$, this is trivial. For part (b), we must show that $RS \supseteq S$. But since rings have unity, we have $RS \supseteq \{1\}S = S$. For part (c), assume that $X \supseteq S$ is a set-theoretic ideal. Then $RX \supseteq RS$. But since X is a set-theoretic ideal, we know by Proposition 1.1.2 that $RX = X$. Hence $X \supseteq RS$ as desired. □

Remark. In plain language, what we've just shown is that an element $a \in R$ is divisible by some element of a set $S \subseteq R$ if and only if a belongs to the set-theoretic ideal generated by S. For example, a polynomial $f \in K[\pi_1]$ has a degree-1 factor if and only if f is an element of the set-theoretic ideal generated by degree-1 polynomials.

The notion of a prime ideal generalizes without effort to set-theoretic ideals.

Definition 1.1.4. Let R denote a ring. Suppose $X \subseteq R$ is a set-theoretic ideal. Then X is said to be *prime* iff for all $a, b \in R$, we have that if $ab \in X$, then $a \in X$ or $b \in X$.

All the examples we've looked at so far are prime. In particular, if ab is a zero-divisor, then a is a zero-divisor or b is a zero-divisor; similarly with the property of being a non-unit. Similarly, the set-theoretic ideal of polynomials having a degree-1 factor is prime; if fg has a degree-1 factor, then either f has a degree-1 factor, or g has a degree-1 factor.

Definition 1.1.5. Let R denote a ring. Suppose $S \subseteq R$ is an arbitrary subset. Then an element $a \in R$ is said to be *squarefree with respect to* S iff for all $x \in S$, we have $x^2 \nmid a$.

Proposition 1.1.6. Given an element $a \in R$ and a subset $S \subseteq R$, the element a is squarefree with respect to S if and only if it is squarefree with respect to the set-theoretic ideal generated by S.

Proof. ($\Rightarrow$). Assume that $a \in R$ is squarefree with respect to S. Consider $y \in RS$. Our goal is to show that $y^2 \nmid a$. We have $y = bx$ for some $a \in R$ and $x \in S$. Assume toward a contradiction that $y^2 \mid a$. Then $b^2x^2 \mid a$. Thus $x^2 \mid a$, a contradiction.

($\Leftarrow$). Assume that $a \in R$ is squarefree with respect to RS. Consider $x \in S$. Our goal is to show that $x^2 \nmid a$. Since $x \in S$, we deduce that $x \in RS$. Thus $x^2 \nmid a$, as desired. □

1.2 Separable polynomials

With the above notions in place, we're ready to consider the notion of a *separable polynomial.* Let K denote a field and consider $f \in K[\pi_1]$. We say that f has *no repeated roots* if and only if it is squarefree with respect to polynomials of degree exactly 1. We say that f has *no repeated factors* if and only if it squarefree with respect to polynomials of degree at least 1. Note that the latter condition is stronger than the former; for example, it's correct to say that $(x^2+1)^2 \in \mathbb{R}[x]$ has no repeated roots, but it's incorrect to say that it has no repeated factors.

Proposition 1.2.1. Over an algebraically closed field, having no repeated roots is equivalent to having no repeated factors.

Proof. It is clear that if a polynomial has no repeated factors, then it has no repeated roots. Let us therefore prove the other direction. Assume toward a contradiction that $f \in \overline{K}[\pi_1]$ has no repeated roots, but does have a repeated factor. Writing $g \in \overline{K}[\pi_1]$ for a repeated factor, we can find $h \in \overline{K}[\pi_1]$ such that $f = g^2 h$. Since g is a (non-trivial) factor, it has degree at least 1. Since f has no repeated roots, this means that g has degree at least 2. Hence by algebraic closedness, we can factor g as a product of degree 1 polynomials, call them A and B. Thus $f = A^2B^2h$. But this means that A is a repeated root of P, a contradiction. □

Definition 1.2.2. Let K denote a field. Then an element $f \in \overline{K}[\pi_1]$ is said to be *separable* if and only if it has no repeated roots, or equivalently, no repeated factors. An element $f \in K[\pi_1]$ is said to be separable if and only if $\overline{f} \in \overline{K}[\pi_1]$ is separable.

Separability is the strongest condition we've considered so far:

$$\text{separable} \ \rightarrow \ \text{no repeated factors} \rightarrow \ \text{no repeated roots.}$$

However, under 'ordinary' circumstances, the left implication in the above chain is actually an equivalence. For example, define $f(x) = x^2 + 1 \in \mathbb{R}[x]$. It's clear that f has no repeated factors. But since $\overline{f}(x)$ as factorizes as $(x + i)(x - i)$, and since $i \neq -i$ in $\mathbb{C}$, hence $\overline{f} \in \mathbb{C}[\pi_1]$. It follows that $f \in \mathbb{R}[\pi_1]$ is separable. Fields like $\mathbb{R}$ in which this equivalence holds are called *perfect.*

Definition 1.2.3. Let K denote a field. Then K is *perfect* if and only if every element of $K[\pi_1]$ with no repeated factors is separable. Otherwise it is *imperfect.*

Examples of perfect fields include every field of characteristic 0 (such as $\mathbb{Q}$ and $\mathbb{R}$, and the p-adic numbers $\mathbb{Q}_p$. Hence every imperfect field is of non-zero characteristic. every finite field (namely $\mathbb{F}_p$ and $\mathbb{F}_{p^n}$). Imperfect fields tend to be harder to find, but the characteristic-2 rational function field $K = \mathbb{F}_2(t)$ is a standard example. To see this, let's take a moment to verify the following:

Proposition 1.2.4. If $K = \mathbb{F}_2(t)$, then the polynomial $P \in K[\pi_1]$ defined by $P(x) = x^2 - t \in K[x]$ is irreducible.

Proof. Assume toward a contradiction that it's reducible, the degree 1 factors being $x - a$ and $x - b$ respectively, with $a, b \in K$ Then

$$x^2 - t = (x - a)(x - b) = x^2 - (a + b)x + ab.$$

Equating coefficients, we see that $b = -a$ and that $ab = 0$. From this we deduce that $a(-a) = 0$. Hence $a = 0$, from which we deduce $b = 0$. Thus $x^2 - t = (x-0)(x-0) = x^2$. Equating coefficients again, we deduce that $-t = 0$, a contradiction. □

Proposition 1.2.5. The field $K = \mathbb{F}_2(t)$ is imperfect.

Proof. We need to find an element of $K[\pi_1]$ that has no repeated factors, but which is not separable. The polynomial $f(x) = x^2 - t \in K[x]$ does the job. Since f irreducible, hence it has no repeated factors. On the other hand, $\overline{P} \in \overline{K}[\pi_1]$ is reducible. In particular, if we let $\sqrt{t}$ denote any square root of t, then we can write $\overline{P}(x) = (x - \sqrt{t})(x + \sqrt{t})$. Since we're in characteristic 2, we deduce that $\overline{P}(x) = (x - \sqrt{t})^2$. Thus $\overline{P} \in \overline{K}[\pi_1]$ is not separable. Thus $P \in K[\pi_1]$ is not separable. Thus $\mathbb{F}_2(t)$ is an imperfect field. □

1.3 Separable elements

So far, we've looked at separable polynomials. This can be used to define the notion of a separable element of a field extension. In what follows, the following fact will be helpful.

Proposition 1.3.1. Let K denote a field. Then for all non-zero polynomials $f, g \in K[\pi_1]$, if $(f) = (g)$, then $\deg(f) = \deg(g)$.

Proof. Since $(f) = (g)$, hence $f \in (g)$, so $g \mid f$. Thus $g \cdot h = f$ for some non-zero $h \in K[\pi_1]$. Thus $\deg(f) = \deg(g) + \deg(h)$. So $\deg(f) \geq \deg(g)$. A similar argument shows the reverse inequality. Hence $\deg(f) = \deg(g)$, as desired. □

Let us now recall the notion of a minimal polynomial. For any field J and any $a \in J$, define $I(a) = \{f \in J[\pi_1] : f(a) = 0\}$. We have:

Definition 1.3.2. Let J/K denote an algebraic field extension. Then the minimal polynomial of an element $a \in J$ is the unique monic polynomial $m_a \in K[\pi_1]$ such that m_a generates the ideal $I(a) \cap K[\pi_1]$ of $K[\pi_1]$.

Proof. Our goal is to show that $I(a) \cap K[\pi_1]$ is generated by a unique monic polynomial in $K[\pi_1]$. To see existence, note that $K[\pi_1]$ is a principal ideal domain, hence the ideal $I(a) \cap K[\pi_1]$ is generated by some $f \in K[\pi_1]$. Since J/K is an algebraic field extension, hence $I(a) \cap K[\pi_1]$ is non-zero, and hence f is non-zero. Thus f has a degree, call it n, and the coefficient c of π_1^n in f is non-zero. Since c is non-zero, it is a unit, so f/c

is a monic polynomial that generates the same ideal as f, namely $I(a) \cap K[\pi_1]$. To see uniqueness, assume that f and g are monic polynomials that generate $I(a) \cap K[\pi_1]$. Our goal is to show that $f = g$. By Proposition 1.3.1, both f and g have the same degree, call it n. Since f and g are both monic, hence the coefficient of π_1^n in both f and g is 1, and this the coefficient of π_1^n in $f - g$ is 0. Now assume toward a contradiction that $f - g$ is non-zero, so that in particular, it has a well-defined degree. Since $\deg(f - g) \leq n$, and since $f - g$ has no degree n term, hence $\deg(f - g) < n$. A routine computation shows that $(f - g)(a) = 0$, and thus $f - g \in I(a) \cap K[\pi_1]$. We deduce that $f \mid f - g$. But this implies that $\deg(f) \leq \deg(f - g)$, and hence that $n \leq \deg(f - g)$, a contradiction. □

Example. The minimal polynomial of $\sqrt{2}$ with respect to $\mathbb{R}/\mathbb{Q}$ is $\pi_1^2 - 2 \in \mathbb{Q}[\pi_1]$, and the minimal polynomial of i with respect to $\mathbb{C}/\mathbb{R}$ is $\pi_1^2 + 1 \in \mathbb{R}[x]$.

Definition 1.3.3. Let J/K denote an algebraic field extension. Then an element $a \in J$ is said to be *separable* (with respect to the extension) if and only if its minimal polynomial $m_a \in K[\pi_1]$ is a separable polynomial. Otherwise we say that a is *inseparable.*

Example. The element $\sqrt{2} \in \mathbb{Q}$ is separable with respect to $\mathbb{R}/\mathbb{Q}$, because the polynomial $m_{\sqrt{2}}(x) = x^2 - 2 \in \mathbb{Q}[x]$ has the property that $\overline{m_{\sqrt{2}}(x)} \in \overline{\mathbb{Q}}[x]$ can be expressed as a product of degree-1 polynomials. In particular:

$$\overline{m_{\sqrt{2}}(x)} = (x - \sqrt{2})(x + \sqrt{2}).$$

Similarly, the element $i \in \mathbb{R}$ is separable with respect to $\mathbb{C}/\mathbb{R}$ because the polynomial $m_i(x) = x^2 + 1 \in \mathbb{Q}[x]$ has the property that $\overline{m_i(x)} \in \overline{\mathbb{Q}}[x]$ can be expressed as a product of degree-1 polynomials. In particular:

$$\overline{m_i(x)} = (x - i)(x + i).$$

On the other hand, let $K = \mathbb{F}_2(t)$ and let $J = \overline{K}$. Then $m_{\sqrt{t}} = x^2 + t$, and we've already seen that $x^2 + t$ is inseparable. Hence $\sqrt{t} \in J$ is inseparable with respect to J/K.

Summarizing what we've done so far, we've defined a notion of *separability* for univariate polynomials over K, which strengthens the condition of having no repeated factors. We've defined the notion of a *perfect field* to exactly mean that this distinction makes no difference. We then used the notion of a separable polynomial to define the notion of a separable element of an algebraic field extension. This can be applied, in particular, to the algebraic field extension $\overline{K}/K$, and we can ask what the separable elements of this field extension are, and in particular whether there's any elements of $\overline{K}$

that fail to separable. If K is perfect, the answer turns out to be 'no', as the following result shows.

Definition 1.3.4. An algebraic field extension J/K is said to be a *separable extension* if and only if every $a \in J$ is separable over the extension, and *purely inseparable* if and only if every $a \in J \setminus K$ is inseparable over the extension.

Remark. The same terminology will be used for morphisms of fields; in particular, a morphism $f : K \to J$ is said to be *separable* if and only if the field extension $J/\mathrm{img}(f)$ is separable, and *purely inseparable* if and only if this extension is purely inseparable.

Proposition 1.3.5. Let K denote a field. Then K is perfect if and only if $\overline{K}/K$ is a separable extension.

Proof. ($\Rightarrow$). Assume K is perfect and consider $a \in \overline{K}$. Our goal is to show that the corresponding minimal polynomial $m_a \in K[\pi_1]$ is separable. Since m_a is a minimal polynomial (this is straightforward to prove), hence in particular m_a has no repeated factors. But since K is perfect, this implies that m_a is separable, as desired.

($\Leftarrow$). Assume $\overline{K}/K$ is not a separable extension. Our goal is to show that K is imperfect. That is, it is enough to find an example of an inseparable polynomial $f \in K[\pi_1]$ such that f has no repeated factors. Since $\overline{K}/K$ is not a separable extension, hence there exists $a \in \overline{K}$ such that m_a is inseparable. But since m_a is irreducible, hence it has no repeated factors. Thus $f := m_a$ provides the desired example. □

Now recall the definition of a Galois extension.

Definition 1.3.6. A *Galois extension* is an algebraic field extension J/K that is normal and separable.

Corollary 1.3.7. Let K denote a field. Then K is perfect if and only if $\overline{K}/K$ is a Galois extension.

This means, among other things, that if K is a perfect field, then infinite Galois theory can be applied to the field extension $\overline{K}/K$. According to Silverman [43, p.1], the assumption that K is a perfect field is unnecessary for much of the theory of elliptic curves, but that also that things are simpler in this case. We'll therefore follow Silverman in studying elliptic curves over perfect fields only, as this prevents us from having to frown at every result we wish to cite to wonder if it really holds in the full generality of K being an arbitrary field. Consequently, we'll make use of the following conventions:

K is a perfect field.

$\overline{K}$ is its algebraic closure

$G_{\overline{K}/K}$ is the Galois group of the field extension $\overline{K}/K$

If more time was available, it would be interesting to explore the question: 'to what extent is the assumption that K is perfect actually necessary for the theory that follows?'

1.4 Degree of a morphism of fields

Degree is usually conceived as a property of field extensions, but it is slightly more convenient for our purposes to define degree as a property of morphisms between fields, as opposed to the more usual definition in which it's a property of field extensions. From this viewpoint, the degree of a morphism of fields measures the extent to which it fails to be surjective.

Definition 1.4.1. If $f : K \to L$ is a morphism of fields, then the *degree* of f, denoted $\deg(f)$ is the dimension of L as a K-module. This is sometimes written $[L : K]$ when the morphism f is clear from the context.

Proposition 1.4.2. If $f : K \to L$ and $g : L \to M$ are morphisms of fields, then $\deg(g \circ f) = \deg(f) \cdot \deg(g)$.

Proof. Our goal is to show that the dimension of M as a K-module is equal to $\deg(f) \cdot \deg(g)$. Let $\varphi : K^{\deg(f)} \to L$ denote a K-linear isomorphism. Let $\psi : L^{\deg(g)} \to M$ denote an L-linear isomorphism. Then the composite

$$K^{\deg(f) \cdot \deg(g)} \to \left(K^{\deg(f)}\right)^{\deg(g)} \xrightarrow{\varphi^{\deg(g)}} L^{\deg(g)} \xrightarrow{\psi} M$$

is a K-linear isomorphism $K^{\deg(f)\cdot\deg(g)} \to M$. Hence the dimension of M as a K-module is $\deg(f) \cdot \deg(g)$, as desired. □

The notion of degree can be decomposed into separable and inseparable parts. In particular, it turns out that the separable elements of a field extension J/K form an intermediate subfield of the extension; the interested reader is directed to Steve Mitchell's notes on the subject [28]. This means, in particular, that given a morphism of fields $f : K \to J$, we can factor f as a composite

$$K \xrightarrow{f} f^{\text{sep}} \hookrightarrow J,$$

where f^{sep} is defined as the intermediate subfield of $J/\text{img}(f)$ consisting of all elements of J that are separable over this extension. This factorization of f into two parts allows us to break up its degree into separable and inseparable parts, as described below.

Definition 1.4.3. If $f : K \to L$ is a morphism of fields, then the *separable degree* of f, denoted $\deg_s(f)$, is the degree of the map $f : K \to f^{\text{sep}}$, and the *inseparable degree* of f, denoted $\deg_i(f)$, is the degree of the inclusion $f^{\text{sep}} \hookrightarrow L$.

Proposition 1.4.4. Let $f : K \to L$ denote a morphism of fields. Then:

$$\deg(f) = \deg_s(f)\deg_i(f).$$

Proof. This is a consequence of applying Proposition 1.4.2 to the following composite:

$$K \xrightarrow{f} f^{\text{sep}} \hookrightarrow L.$$ □

An easy but helpful observation observation is that f is separable if and only if $\deg_s(f) = \deg(f)$.

1.5 Transcendence degree

Recall that given a morphism of fields f, the degree of f measures the extent to which f fails to be surjective. Unfortunately, the notion of degree can sometimes gives uninformative results. For instance, the inclusions $\mathbb{R} \to \mathbb{R}(x)$ and $\mathbb{R} \to \mathbb{R}(x, y)$ both have degree $\aleph_0$, yet clearly the former 'fails to be surjective' to a much greater extent than the latter. In situations such as this, the notion of *transcendence degree* can be useful. The material in this section is mainly from Alex Wright's notes on the subject [46].

Definition 1.5.1. Given a K-ring R, a subset $\{r_1, \cdots, r_n\}$ of R is said to be algebraically independent over K if and only if for all $f \in K[\pi_1, \ldots, \pi_n]$, we have

$$f(r_1, \ldots, r_n) = 0 \to f = 0.$$

Given a morphism of fields $\varphi : K \to L$, a subset $\{l_1, \ldots, l_n\}$ of φ is said to be algebraically independent over φ if and only if it is algebraically independent over K when L is made into a K-algebra via φ.

Proposition 1.5.2. Every K-ring has a maximal subset of elements that are algebraically independent over K.

Proof. This is a straightforward application of Zorn's Lemma. The details are left to the reader. □

Definition 1.5.3. Let $\varphi : K \to L$ denote a morphism of fields. If L has a maximal algebraically independent subset over φ that is finite with n elements, then every maximal algebraically independent subset over φ has exactly n elements, and we refer to n as the transcendence degree of φ.

Remark. The degree of a morphism of fields should probably have been called its *linear degree*, and the transcendence degree should probably have been called its *algebraic degree.* It's a bit late to change things now, but being aware of this analogy can aid the memory a bit. It's also helpful when talking to other mathematicians; 'no sorry, I meant its *linear* degree' instantly conveys the intended meaning.

Proof. This is proven in Alex Wright's notes [46, p.4]. □

Corollary 1.5.4. The transcendence degree of the inclusion of K into $K(\pi_1, \ldots, \pi_n)$ is exactly n.

Chapter 2

Closure operators and Galois connections

The material in this section is mainly standard results, but it's also interspersed with my own thinking about the Galois connection (I, V) at the heart of classical algebraic geometry, and the closure operators it induces. Fundamentally, it is general facts about Galois connections that allow the classical theory of the Zariski topology to function, and this point of view has influenced how I look at Galois connections in general. Consequently there is no one source where most of this material comes from. However, the interested reader is directed to the article *A primer on Galois connections* [18] where I learned much of the material. Davey and Priestley's textbook on lattices and order theory [15] was also very helpful, and contains, among other things, the relevant lattice-theoretic preliminaries.

2.1 Closure Operators

The notion of closure is ubiquitous in mathematics, and especially helpful in the context of algebraic geometry, where several different notions of closure are often in play at the same time. For this reason, it will be helpful to have a general framework available.

Definition 2.1.1. Given a poset P and a function $f : P \to P$, we make the following definitions.

(a) Elements $x \in P$ satisfying $f(x) = x$ are called *fixed points* of f. We write

$$\operatorname{fix}(f) = \{x \in P : f(x) = x\}.$$

(b) Elements $x \in P$ satisfying $f(x) \leq x$ are said to be *closed* with respect to f. We write
$$\mathcal{C}(f) = \{x \in P : f(x) \leq x\}.$$

(c) The function f is said to be *inflative* if and only if for all $x \in P$ we have $f(x) \geq x$.

(d) The function f is said to be *idempotent* if and only if for all $x \in P$ we have $f(f(x)) = f(x)$, which can be written more tersely as $\mathrm{img}(f) \subseteq \mathrm{fix}(f)$.

(e) The function f is said to be *sub-idempotent* if and only if for all $x \in P$ we have $f(f(x)) \leq f(x)$. More tersely, this can be written $\mathrm{img}(f) \subseteq \mathcal{C}(f)$.

It's clear that $\mathrm{fix}(f) \subseteq \mathcal{C}(f)$. The following is a partial converse to this statement.

Proposition 2.1.2. If $f : P \to P$ is an inflative function, then $\mathcal{C}(f) \subseteq \mathrm{fix}(f)$.

Proof. Assume f is inflative and consider $x \in \mathcal{C}(f)$. Our goal is to show that $x \in \mathrm{fix}(f)$. Since $x \in \mathcal{C}(f)$, hence $f(x) \leq x$. Since f is inflative, hence $x \leq f(x)$. Hence $x = f(x)$. We deduce that $x \in \mathrm{fix}(f)$, as desired. □

Every idempotent function is clearly sub-idempotent. The next result is a partial converse to this statement.

Proposition 2.1.3. Suppose $f : P \to P$ is inflative and monotone. Then if f is sub-idempotent, then f is idempotent.

Proof. Consider $x \in P$. Our goal is to show that $f(f(x)) = f(x)$. Since f is sub-idempotent, we have $f(f(x)) \leq f(x)$. Since f is inflative, we have $x \leq f(x)$. Since f is monotone, we deduce that $f(x) \leq f(f(x))$. Thus $f(f(x)) = f(x)$, as desired. □

Definition 2.1.4. Let $f : P \to P$ denote a monotone mapping. Then the following are equivalent:

(a) f is inflative and sub-idempotent

(b) f is inflative and idempotent

(c) For all $x, y \in P$ we have $x \leq f(y) \iff f(x) \leq f(y)$.

We call a monotone mapping satisfying any (and hence all) of these conditions a *closure operator* on X.

Proof. The equivalence between the first two conditions follows from Proposition 2.1.3. Hence we must show that the property defined by the first two conditions both implies and is implied by the last condition. For the ($\Leftarrow$) direction, observe that by taking $y := x$ in (b) we obtain $x \leq \mathrm{cl}(x)$ and thus deduce that cl is inflative. Taking $x := \mathrm{cl}(y)$ in (b) we obtain $\mathrm{cl}(\mathrm{cl}(y)) \leq \mathrm{cl}(y)$ and thus deduce that cl is sub-idempotent.

For the ($\Rightarrow$) direction, there are two parts. For the first part, assume $x \leq \mathrm{cl}(y)$. Our goal is to show that $\mathrm{cl}(x) \leq \mathrm{cl}(y)$. Applying cl to both sides of our premise, we deduce $\mathrm{cl}(x) \leq \mathrm{cl}(\mathrm{cl}(y))$. By sub-idempotency, we infer that $\mathrm{cl}(x) \leq \mathrm{cl}(y)$, as desired. For the second part, assume $\mathrm{cl}(x) \leq \mathrm{cl}(y)$. Our goal is to show that $x \leq \mathrm{cl}(y)$. But from inflativeness, we have $x \leq \mathrm{cl}(x)$, and by transitivity the desired result follows. □

There are many closure operators of relevance to the study of a ring R. One of the most basic is the closure operator $\mathcal{P}(R) \to \mathcal{P}(R)$ that assigns to each $X \in \mathcal{P}(R)$ the smallest subring that includes X. Even more fundamental is the closure operator that assigns to each $X \in \mathcal{P}(R)$ the smallest ideal that includes X.

The above examples suggest a general way of obtaining closure operators. In particular, the closure of a thing with respect to a desirable property can usually be defined as the smallest thing satisfying the desired property that includes the entity of interest. This way of thinking leads to a completely general characterization of closure operators that we now describe.

Definition 2.1.5. Given a poset P and a subset $S \subseteq P$, define that S *induces a closure operator* if and only if for all $p \in P$, there is a smallest $s \in S$ such that $p \leq s$. In this case, we write $\mathrm{cl}_S(p)$ for the smallest such s. We refer to cl_S as the *closure operator induced by* S.

It remains to be shown that this is indeed a closure operator.

Proposition 2.1.6. For all $s \in S$, we have $\mathrm{cl}_S(s) = s$.

Proof. This follows immediately from the definition. □

Proposition 2.1.7. The function $\mathrm{cl}_S : P \to P$ defined above is indeed a closure operator.

Proof. Monotonicity is clear. For inflativeness, note that since $\mathrm{cl}_S(x)$ is the smallest $s \in S$ with $x \leq s$, hence it's clear that $x \leq \mathrm{cl}_S(x)$. It remains to show that $\mathrm{cl}_S(\mathrm{cl}_S(x)) \leq \mathrm{cl}_S(x)$. But since $\mathrm{cl}_S(x) \in S$, this follows from Proposition 2.1.6. □

We can also go the other way, and obtain from any closure operator $\mathrm{cl} : P \to P$ a corresponding subset of P. Indeed, these are inverse processes, as the next Proposition shows.

Proposition 2.1.8. Let P denote a poset. Then:

(a) For all $S \in \mathcal{P}(P)$, if S induces a closure operator, then $\mathcal{C}(\mathrm{cl}_S) = S$.

(b) For all closure operators $\mathrm{cl} : P \to P$, we have $\mathrm{cl}_{\mathcal{C}(\mathrm{cl})} = \mathrm{cl}$.

Proof. Part (a). Suppose $S \in \mathcal{P}(P)$ induces a closure operator. Our goal is to show that for all $x \in P$, we have

$$x \in S \iff \mathrm{cl}_S(x) \leq x.$$

For the forward direction, assume $x \in S$. Then by Proposition 2.1.6, we have $\mathrm{cl}_S(x) = x$. So the weaker statement $\mathrm{cl}_S(x) \leq x$ is immediate. For the backward direction, assume $\mathrm{cl}_S(x) \leq x$. We wish to show that $x \in S$. Suppose toward a contradiction that $x \notin S$. Then since $\mathrm{cl}_S(x) \in S$, we deduce that $\mathrm{cl}_S(x) \neq x$. Since cl_S is a closure operator, this implies $\mathrm{cl}_S(x) > x$. But this contradicts our hypothesis that $\mathrm{cl}_S(x) \leq x$.

Part (b). Consider $p \in P$. We must show that $\mathrm{cl}_{\mathcal{C}(\mathrm{cl})}(p) = \mathrm{cl}(p)$. There are two directions. To show that $\mathrm{cl}_{\mathcal{C}(\mathrm{cl})}(p) \leq \mathrm{cl}(p)$, it suffices to show that $\mathrm{cl}_{\mathcal{C}(\mathrm{cl})}(\mathrm{cl}(p)) \leq \mathrm{cl}(p)$. But by Proposition 2.1.6 is immediate. For the other direction, we must show that $\mathrm{cl}_{\mathcal{C}(\mathrm{cl})}(p) \geq \mathrm{cl}(p)$. Assume $q \geq p$ and $q \in \mathcal{C}(\mathrm{cl})$. Our goal is to show that $q \geq \mathrm{cl}(p)$. Since $q \in \mathcal{C}(\mathrm{cl})$, we have $q \geq \mathrm{cl}(q)$. Since $q \geq p$, we have $\mathrm{cl}(q) \geq \mathrm{cl}(p)$. Putting these two inequalities together, we deduce $q \geq \mathrm{cl}(p)$, as required. □

We have established a bijective correspondence between closure operators on a poset P and those subsets of P that induce closure operators. But recall that to prove that each subset S of a ring R generates a subring, we usually take the intersection of all subrings that include S, in order to establish existence. A similar technique is used in topology; to show that every subset of a topological space has a closure, we take the intersection of the closed sets that include it. This kind of thinking can be generalized to yield an improved characterization of closure operators in the special case where P is a complete lattice. The following lemma is helpful in this regard.

Proposition 2.1.9. Let $f : P \to Q$ denote an order-preserving map between complete lattices. Then for all sets I and all functions $I \xrightarrow{\varphi} P$, we have

$$f\left(\bigwedge_{i\in I} \varphi_i\right) \leq \bigwedge_{i\in I} f(\varphi_i).$$

Proof. Since $\forall j \in I : \bigwedge_{i\in I} \varphi_i \leq \varphi_j$, we deduce $\forall j \in I : f\left(\bigwedge_{i\in I} \varphi_i\right) \leq f(\varphi_j)$. It follows that $f\left(\bigwedge_{i\in I} \varphi_i\right) \leq \bigwedge_{j\in I} f(\varphi_j)$, as desired. □

Proposition 2.1.10. Let P denote a complete lattice and $S \subseteq P$ denote a subset thereof. Then S induces a closure operator if and only if S is closed under arbitrary meets.

Proof. ($\Rightarrow$). Suppose S induces a closure operator. Consider a set I and a function $I \xrightarrow{\varphi} S$. Our goal is to show that $\bigwedge_{i\in I} \varphi_i \in S$, where the meet above is taken in P. Since $\mathcal{C}(\mathrm{cl}_S) = S$, it suffices to show that $\bigwedge_{i\in I} \varphi_i \in \mathcal{C}(\mathrm{cl}_S)$. Thus our goal is to show that $\mathrm{cl}_S\left(\bigwedge_{i\in I} \varphi_i\right) \leq \bigwedge_{i\in I} \varphi_i$. By Proposition 2.1.9, it suffices to show that $\bigwedge_{i\in I} \mathrm{cl}_S(\varphi_i) \leq \bigwedge_{i\in I} \varphi_i$. But since $\forall \in I : \varphi_i \in S$, we have $\forall i \in I : \mathrm{cl}_S(\varphi_i) = \varphi_i$ by Proposition 2.1.6 above. This proves the above inequality.

($\Leftarrow$). Suppose S is closed under meets. Our goal is to show that S induces a closure operator. Consider $p \in P$. We seek to show that there is a smallest $s \in S$ such that $p \leq s$. Define $s := \bigwedge_{q\in S:q\geq p} q$. We must prove firstly that $p \leq s$, and secondly that $p \leq t$ and $t \in S$ imply $s \leq t$. To see that $p \leq s$, we make the following computation:

$$\begin{aligned} & p \leq s \\ & \iff p \leq \bigwedge_{q\in S:q\geq p} q \\ & \iff \forall q \in S(q \geq p \to p \leq q) \\ & \iff \text{True} \end{aligned}$$

So assume $p \leq t$ and $t \in S$. We must show that $s \leq t$. That is, we must show that $\bigwedge_{q\in S:q\geq p} \leq t$. That is, we must show that $t \geq p$ and that $t \in S$. But these are precisely our hypotheses. □

2.2 Galois connections

There is another, quite different way of obtaining closure operators with special significance for algebraic geometry. In particular, closure operators arise naturally from Galois connections, defined below:

Definition 2.2.1. Given posets X and Y, a *Galois connection* (f, g) from X to Y consists of order-reversing maps $f : X \to Y$ and $g : Y \to X$ such that either and hence both of the following conditions hold:

(a) $f(x) \geq y \iff x \leq g(y)$

(b) $y \leq f(g(y))$ and $x \leq g(f(x))$

Proof. ($\Rightarrow$). We may take $x := g(y)$ in (a). This yields $f(g(y)) \geq y \iff g(y) \geq g(y)$. We thus conclude that $y \leq f(g(y))$. Similarly, we may take $y := f(x)$, thereby obtaining $x \leq g(f(x))$.

($\Leftarrow$). Assume $f(x) \geq y$. Our goal is to show that $x \leq g(y)$. Applying the order-reversing function g to both sides of our premise, we deduce $g(f(x)) \leq g(y)$. Using (b) we obtain $x \leq g(y)$, as desired. The other direction is similar. □

The following definition gives us a general mechanism for obtaining Galois connections.

Definition 2.2.2. Given sets S and T and a relation $R \in \mathcal{P}(S \times T)$, we obtain a Galois connection

$$(R_\forall, R^\forall) : \mathcal{P}(S) \to \mathcal{P}(T)$$

given as follows:

$$R_\forall(X) = \{y \in T : \forall x \in X, R(x,y)\}, \qquad R^\forall(Y) = \{x \in S : \forall y \in Y, R(x,y)\}.$$

Proof. The first part is to show that $R_\forall$ and $R^\forall$ are order-reversing. We'll prove the former, as the latter is similar. Consider $X_0, X_1 \in \mathcal{P}(S)$ satisfying $X_0 \subseteq X_1$. Our goal is to show that $R_\forall(X_0) \supseteq R_\forall(X_1)$. So consider $y \in R_\forall(X_1)$. We must show that $y \in R_\forall(X_0)$. That is, we must show that $\forall x \in X_0, R(x,y)$. Since $X_0 \subseteq X_1$, it suffices to show that $\forall x \in X_1, R(x,y)$. But this is precisely the statement that $y \in R_\forall(X_1)$, which is true by hypothesis. We deduce that $R_\forall$ is order-reversing. In a similar way, one can check that $R^\forall$ is order-reversing.

Let us now verify (a) from the definition of a Galois connection. Focusing on the left-hand side, we obtain:

$$\begin{aligned} & R_\forall(X) \supseteq Y \\ \iff & \{y \in T : \forall x \in X, R(x,y)\} \supseteq Y \\ \iff & \forall y \in Y, \forall x \in X, R(x,y) \end{aligned}$$

Focusing on the right-hand side, we obtain:

$$\begin{aligned} & X \subseteq R_\forall(Y) \\ \iff & X \subseteq \{x \in S : \forall y \in Y, R(x,y)\} \\ \iff & \forall x \in X, \forall y \in Y, R(x,y) \end{aligned}$$

But as these are equivalent, this completes the proof. □

Let us now consider some examples. Let E/F denote a field extension. An *automorphism* of E/F is a field automorphism φ of E such that for all $x \in F$ we have $\varphi(x) = x$. Recall that the group of field automorphisms of E/F is called its *Galois group*, denoted $G_{E/F}$. Define a relation as follows:

$$\mathrm{fix} \in \mathcal{P}(G_{E/F} \times E)$$

$$\mathrm{fix}(\varphi, x) \iff \varphi(x) = x.$$

The forward map $\mathrm{fix}_\forall : \mathcal{P}(G_{E/F}) \to \mathcal{P}(E)$ turns a set of automorphisms of E/F into its set of fixed points and is usually denoted $E^H := \mathrm{fix}_\forall(H)$. The backward map $\mathrm{fix}^\forall : \mathcal{P}(E) \to \mathcal{P}(G_{E/F})$ turns a set of elements of E into the set of automorphisms of E/F that fixes every element; there is no special notation for this.

Proposition 2.2.3 (Fundamental Theorem of Galois Theory, General Version)**.** If a finite field extension E/F is normal and separable, then for all $H \subseteq G_{E/F}$, we have that $(\mathrm{fix}^\forall \circ \mathrm{fix}_\forall)(H)$ is the smallest group that includes H, and for all sets $S \subseteq E$, we have that $(\mathrm{fix}_\forall \circ \mathrm{fix}^\forall)(S)$ is the smallest intermediate field of E/F that includes S.

We usually restrict the domains and codomains of the aforementioned functions so that $\mathcal{P}(E)$ is replaced by the poset $\mathrm{Sub}(E/F)$ of intermediate fields of E/F and $\mathcal{P}(G_{E/F})$ is replaced by the poset $\mathrm{Sub}(G_{E/F})$ of subgroups of $G_{E/F}$. Write $\mathrm{f}^\forall$ and $\mathrm{f}_\forall$ for the restricted maps. We obtain a Galois connection

$$(\mathrm{f}_\forall, \mathrm{f}^\forall) : \mathrm{Sub}(G_{E/F}) \to \mathrm{Sub}(E/F),$$

and a more traditional way to phrase the fundamental theorem of Galois theory is to refer only to $(\mathrm{f}_\forall, \mathrm{f}^\forall)$. For example:

Proposition 2.2.4 (Fundamental Theorem of Galois Theory, Special Version)**.** If a finite field extension E/F is normal and separable, then for all $H \subseteq G_{E/F}$, we have that $\mathrm{f}^\forall$ and $\mathrm{f}_\forall$ are inverse functions.

Another good example of a Galois connection, which is fundamental to classical algebraic geometry, is obtained as follows: To each natural number n, let us assign a relation given as follows:

$$R(n) \in \mathcal{P}(K[\pi_1, \ldots, \pi_n] \times \overline{K}^n)$$

$$R(n) = \{(f, P) : f(P) = 0\}.$$

Classical algebraic geometry can largely be defined as the study the Galois connection induced by each $R(n)$. By convention, the corresponding forward map is denoted $V^n := R(n)_\forall$. This notation is chosen because V^n turns a set of polynomials (in n-many variables) into the set of all points in $\overline{K}^n$ causing those polynomials to Vanish. An explicit definition is given below:

$$\begin{aligned} \mathcal{P}(K[\pi_1, \ldots, \pi_n]) &\xrightarrow{V^n} \mathcal{P}(\overline{K}^n) \\ X &\longmapsto \{P \in \overline{K}^n : \forall f \in X, f(P) = 0\}. \end{aligned}$$

The backward map is denoted $I^n := R(n)^\forall$. It's chosen because I^n turns a set of points (in n-dimensional space) into the Ideal of all polynomials that vanish on that set:

$$\begin{aligned} \mathcal{P}(\overline{K}^n) &\xrightarrow{I^n} \mathcal{P}(K[\pi_1, \ldots, \pi_n]) \\ Y &\longmapsto \{f \in K[\pi_1, \ldots, \pi_n] : \forall P \in Y, f(P) = 0\}. \end{aligned}$$

The proof that $I^n(Y)$ is always an ideal, for any set $Y \subseteq \overline{K}^n$, will come later. For now, let us take these two examples as motivation for developing some general results about Galois connections. For the rest of this subsection, let X and Y denote posets and $(f, g) : X \to Y$ denote a Galois connection.

Proposition 2.2.5 (Ternary Law)**.** We have:

$$f \circ g \circ f = f, \qquad g \circ f \circ g = g.$$

Proof. Let us verify the first identity, as the verification of the second is similar. Using definition (b) of a Galois connection, we know that $g \circ f \geq \mathrm{id}$, and post-composing by f (which is order-reversing) this tells us that $f \circ g \circ f \leq f \circ \mathrm{id}$, which implies $f \circ g \circ f \leq f$, as desired. On the other hand, from (b) we also know that $f \circ g \geq \mathrm{id}$, and pre-composing by f tells us that $f \circ g \circ f \geq \mathrm{id} \circ f$, which implies that $f \circ g \circ f \geq f$. Taking these results together, we infer that $f \circ g \circ f = f$, which was our goal. □

Proposition 2.2.6. Let X and Y denote posets and $(f, g) : X \to Y$ denote a Galois connection. Then $g \circ f$ is a closure operator on X and $f \circ g$ is a closure operator on Y.

Proof. Let's show that $g \circ f$ is a closure operator on X, since the proof that $f \circ g$ is a closure operator on Y is similar. We must show that $g \circ f$ is order-preserving, inflative and sub-idempotent. Observe that since f and g and order-reversing, hence $g \circ f$ is order-preserving. Observe also that $g \circ f$ is inflative directly from (b) in the definition of a Galois connection. Finally note that from Proposition 2.2.5, we know that $f \circ g \circ f = f$, and hence that $g \circ f \circ g \circ f = g \circ f$. This shows that $g \circ f$ is idempotent, and hence sub-idempotent. $\square$

Definition 2.2.7. An element $x \in X$ is said to be *closed* if and only if $x \in \text{img}(g)$. Write $\mathcal{C}(X)$ for the set of closed elements of X.

Proposition 2.2.8. We have $\mathcal{C}(X) = \mathcal{C}(g \circ f)$.

Proof. For the $(\subseteq)$ direction, assume $x \in \mathcal{C}(X)$. Then $x \in \text{img}(g)$. So we can find $y \in Y$ with $x = g(y)$. Thus $g(f(x)) = g(f(g(y)))$. From the Proposition 2.2.5, we deduce $g(f(x)) = g(y)$. Hence $g(f(x)) = x$. So $x \in \mathcal{C}(g \circ f)$. For the $(\supseteq)$ direction, assume $x \in \mathcal{C}(g \circ f)$. Our goal is to show that $x \in \mathcal{C}(g)$. Hence $x \in \text{fix}(g \circ f)$ by Propositions 2.2.6 and 2.1.2. Thus $x \in \text{img}(g \circ f)$. Ergo $x \in \text{img}(g)$, as desired. $\square$

Proposition 2.2.9. The closed elements of X induce a closure operator, and the induced closure operator is $g \circ f$.

Proof. Since $g \circ f$ is a closure operator by Proposition 2.2.6, we deduce that $\mathcal{C}(g \circ f)$ induces a closure operator. Thus $\mathcal{C}(X)$ induces a closure operator, by Proposition 2.2.8, from which it also follows that $\text{cl}_{\mathcal{C}(X)} = \text{cl}_{\mathcal{C}(g \circ f)}$. Hence from Proposition 2.1.8 we have $\text{cl}_{\mathcal{C}(X)} = g \circ f$. $\square$

Proposition 2.2.10. If X is a complete lattice, then an arbitrary meet of closed elements of X is itself closed.

Proof. This is a consequence of the fact that $\mathcal{C}(X)$ induces a closure operator by Proposition 2.1.8, and therefore Proposition 2.1.10 applies. $\square$

Proposition 2.2.11. Let X and Y denote complete lattice and $(f, g) : X \to Y$ denote a Galois connection. Suppose I is a set and $I \xrightarrow{x} X$ is a function. Then

$$f\left(\bigvee_{i \in I} x_i\right) = \bigwedge_{i \in I} f(x_i).$$

Proof. Consider $y \in Y$. We compute:

$$\begin{aligned} y \le f\left(\bigvee_{i\in I} x_i\right) &\iff g(y) \ge \bigvee_{i\in I} x_i \\ &\iff \forall(i \in I)\ g(y) \ge x_i \\ &\iff \forall(i \in I)\ y \le f(x_i) \\ &\iff y \le \bigwedge_{i\in I} f(x_i) \end{aligned}$$

This completes the proof.

□

Chapter 3

Curves and varieties

The material in this chapter mainly comes from Silverman's book on elliptic curves [43], Milne's notes on algebraic geometry [27], and Ash and Gross's pedagogical book on the notions of degree and multiplicity [1].

3.1 Affine space and the Zariski topology

Recall that given a collection of polynomials $X \subseteq K[\pi_1, \ldots, \pi_n]$, we define the *vanishing set of* X as follows:

$$V_K^n(\mathcal{F}) = \{P \in \overline{K}^n : \forall f \in \mathcal{F} : f(P) = 0\}.$$

Recall also that given a collection of points $Y \subseteq \overline{K}^n$, we define the ideal of polynomials that vanish on Y as follows:

$$I_K^n(Y) = \{f \in K[\pi_1, \ldots, \pi_n] : \forall P \in Y : f(P) = 0\}.$$

We'll usually drop the subscripts (and sometimes superscripts) for brevity. Most of the basic results about V^n and I^n are special cases of the fact that the pair

$$(I^n, V^n) : \overline{K}^n \to K[\pi_1, \ldots, \pi_n]$$

is a Galois connection. In particular, from Proposition 2.2.5 we deduce:

$$V^n \circ I^n \circ V^n = V^n \qquad I^n \circ V^n \circ I^n = I^n.$$

And from Proposition 2.2.11, we deduce

$$V^n\left(\bigcup_{i\in I}\mathcal{F}_i\right)=\bigcap_{i\in I}V(\mathcal{F}_i),\qquad I^n\left(\bigcup_{i\in I}X_i\right)=\bigcap_{i\in I}I^n(X_i)$$

Some further facts that do not follow from the general theory of Galois connections will now be demonstrated. They're significant because they'll allow us to obtain a topology on $\overline{K}^n$.

Proposition 3.1.1. Defining $R := K[\pi_1, \dots, \pi_n]$, we have that $V^n(R) = \emptyset$.

Proof. Assume toward a contradiction that $P \in V(R)$. Since $R \supseteq \{1\}$, thus $V(R) \subseteq V(\{1\})$. So $P \in V(\{1\})$. We deduce that $1(P) = 0$, from which it follows that $1 = 0$, a contradiction. □

Definition 3.1.2. Given sets $\mathcal{F}, \mathcal{G} \subseteq K[x_1, \dots, x_n]$, define:

$$\mathcal{F}\mathcal{G} := \{fg : f \in \mathcal{F}, g \in \mathcal{G}\}.$$

Proposition 3.1.3. If $\mathcal{F}, \mathcal{G} \subseteq K[x_1, \dots, x_n]$ are collections of polynomials, then:

$$V(\mathcal{F}\mathcal{G}) = V(\mathcal{F}) \cup V(\mathcal{G}).$$

Proof. ($\supseteq$). Consider $x \in V(\mathcal{F}) \cup V(\mathcal{G})$. Our goal is to show that $x \in V(\mathcal{F}\mathcal{G})$. So consider $h \in \mathcal{F}\mathcal{G}$. Our goal is to show that $h(x) = 0$. Since $h \in \mathcal{F}\mathcal{G}$, we know that $h = fg$ for some $f \in \mathcal{F}$ and $g \in \mathcal{G}$. Since $x \in V(\mathcal{F}) \cup V(\mathcal{G})$, there are two cases. In the case $x \in V(\mathcal{F})$, we have $f(x) = 0$ and hence $f(x)g(x) = 0$. Thus $h(x) = 0$. The case where $x \in V(\mathcal{G})$ is similar.

($\subseteq$). We'll show the contrapositive. Suppose $x \notin V(\mathcal{F}) \cup V(\mathcal{G})$. Our goal is to show that $x \notin V(\mathcal{F}\mathcal{G})$. So consider $h \in \mathcal{F}\mathcal{G}$. Our goal is to show that $h(x) \neq 0$. Since $h \in \mathcal{F}\mathcal{G}$, we know that $h = fg$ for some $f \in \mathcal{F}$ and $g \in \mathcal{G}$. It suffices to show that $f(x)g(x) \neq 0$. Since $x \notin V(\mathcal{F}) \cup V(\mathcal{G})$, thus $x \notin V(\mathcal{F})$ and $x \notin V(\mathcal{G})$. Since $x \notin V(\mathcal{F})$ thus $f(x) \neq 0$. Since $x \notin V(\mathcal{G})$ thus $g(x) \neq 0$. We deduce that $f(x)g(x) \neq 0$, as desired. □

Definition 3.1.4. Given a natural number n, we write $\mathbb{A}^n_K$ for the set $\overline{K}^n$. We make this into a topological space by declaring that $A \subseteq \mathbb{A}^n_K$ is (Zariski) *closed* if and only if $A \in \operatorname{img}(V^n_K)$.

Remark. The set $\mathbb{A}^n_K$ is called affine space over K and the topology defined above is called the Zariski topology on affine space.

Proof. We need to show that the closed sets endow $\mathbb{A}_K^n$ with a closed topology. Since $\mathcal{P}(\mathbb{A}_K^n)$ is a complete lattice in which meets are intersections, hence by Proposition 2.1.10 we deduce that an arbitrary intersection of Zariski-closed subsets of $\mathbb{A}_K^n$ is Zariski-closed. This also implies that $\mathbb{A}_K^n$ is Zariski-closed in itself, due to the standard convention in lattice theory that the empty meet is the maximum element of the ambient poset. Hence it remains to to show (a) that $\emptyset$ is Zariski-closed, and (b) that the union of two Zariski-closed sets is Zariski-closed.

For part (a), it suffices to show that $\emptyset = V(K[\pi_1, \ldots, \pi_n])$. But this is Proposition 3.1.1. For part (b), suppose we're given Zariski-closed sets $A, B \subseteq \overline{K}^n$. We want to show that $A \cup B$ is Zariski-closed. We know that $A = V^n(\mathcal{F})$ and $B = V^n(\mathcal{G})$ for some $\mathcal{F}, \mathcal{G} \subseteq K[\pi_1, \ldots, \pi_n]$. Thus we wish to show that $V(\mathcal{F}) \cup V(\mathcal{G})$ is Zariski-closed. Hence it suffices to show that $V(\mathcal{F}) \cup V(\mathcal{G}) = V(\mathcal{F}\mathcal{G})$. But this is the result of Proposition 3.1.3. □

Proposition 3.1.5. Given $A \subseteq \mathbb{A}_K^n$, the set A is closed if and only if $V(I(A)) = A$.

Proof. This is a consequence of Proposition 2.2.9. □

Having looked at the function V that builds vanishing sets, let us now take a look at the function I which builds ideals. The following is trivial, but we include it for the sake of completeness.

Proposition 3.1.6. For all $A \subseteq \mathbb{A}_K^n$, the set $I(A) \subseteq K[\pi_1, \ldots, \pi_n]$ is a radical ideal.

Proof. We must show the following:

(a) If $f, g \in I(A)$, then $f - g \in I(A)$.

(b) If $a \in K$ and $f \in I(A)$, then $af \in I(A)$.

(c) If $f^n \in I(A)$ then $f \in I(A)$.

For (a), assume $f, g \in I(A)$. Consider $P \in A$. Our goal is to show that $(f-g)(P) = 0$. It's enough to show that $f(P) = 0$ and that $g(P) = 0$. But these are true by hypothesis. For (b), assume $f \in I(A)$. Our goal is to show that $af \in I(A)$. So consider $P \in A$. Our goal is to show that $(af)(P) = 0$. It is enough to show that $f(P) = 0$. But this is true by hypothesis. For (c), assume $f^n \in I(A)$. Our goal is to show that $f \in I(A)$. Consider $P \in A$. It is enough to show that $f(P) = 0$ We know that $(f(P))^n = 0$. But since K is a field, it is an integral domain, hence a reduced ring, and hence has no nilpotent elements. We deduce that $f(P) = 0$, as desired. □

The converse to the above result is also true, but much harder to prove.

Proposition 3.1.7 (Semirational Nullstellensatz)**.** For all ideals $\mathfrak{a} \subseteq K[\pi_1, \ldots, \pi_n]$, we have $I(V(\mathfrak{a})) = \sqrt{\mathfrak{a}}$.

Proof. This is Theorem 11.9 in Pete L. Clarke's notes [11, p.210]. □

Remark. The above theorem would fail if we were to look for solutions in K^n instead of $\overline{K}^n$. For example, define $f \in \mathbb{R}[\pi_1]$ by writing $f(x) = x^2 + 1.$. Then $V(f) \cap \mathbb{R}$ is empty. Hence $I(V(f) \cap \mathbb{R})$ is $\mathbb{R}[x]$, but $\sqrt{(f)}$ is (f).

Proposition 3.1.8. For all radical ideals $\mathfrak{a} \subseteq K[\pi_1, \ldots, \pi_n]$, there exists $A \subseteq \mathbb{A}^n_K$ such that the set $I(A) = \mathfrak{a}$.

Proof. Define $A = V(\mathfrak{a})$. Our goal is to show that $I(V(\mathfrak{a})) = \mathfrak{a}$. By Proposition 3.1.7, it is enough to show that $\sqrt{\mathfrak{a}} = \mathfrak{a}$. But since $\mathfrak{a}$ is a radical ideal, this follows from Proposition 2.1.6. □

We've seen abstractly that radical ideals and Zariski-closed sets are somehow the same, but let's take a moment to look at this correspondence more concretely. Consider the polynomial $f \in \mathbb{R}[\pi_1, \pi_2]$ given by $f(x, y) = x^2 - y^2$. To see that this is a radical ideal, assume toward a contradiction that there exist $a, b, c \in \mathbb{R}$ such that $f = (ax + by + c)^2$. Equating coefficient, we see immediately that $a^2 = 1$ and that $b^2 = 1$ and that $c^2 = 0$. Hence $f = (x + y)^2$ or $f = (x - y)^2$, and both possibilities lead to a contradiction. Hence it's a radical ideal.

Let's now compute the vanishing set, writing $\{\Phi\}$ as shorthand for $\{(x, y) \in \mathbb{C}^2 : \Phi\}$.

$$\begin{aligned} V(f) &= \{x^2 - y^2 = 0\} \\ &= \{(x - y)(x + y) = 0\} \\ &= \{x - y = 0\} \cup \{x + y = 0\}. \end{aligned}$$

So the radical ideal $(\pi_1^2 - \pi_2^2)$ corresponds to two lines crossing at the origin under the above correspondence. The two lines are in some sense 'irreducible' pieces, where irreducibility ought to be a condition like connectedness, but stronger (since the aforementioned $V(f)$ consisting of two lines crossing ought not to be irreducible). The question naturally arises of whether there's a good definition of irreducibility in this context, and if so, whether any Zariski-closed set can be decomposed as a union of irreducible piece like this. The answer to both questions is 'yes.'

3.2 Noetherian spaces and Irreducibility

Definition 3.2.1. Given a topological space X, we say that X is *irreducible* if and only if it is non-empty, and for all proper closed sets $A, B \subseteq X$, from $X = A \cup B$ we can infer $X = A$ or $X = B$.

Proposition 3.2.2. Given a Zariski-closed subset A of $\mathbb{A}^n_K$, the following are equivalent:

1. A is irreducible in the Zariski topology
2. $A = V(\mathfrak{p})$ for some prime ideal $\mathfrak{p}$
3. $I(A)$ is a prime ideal

Proof. (1) ⇒ (3). Assume A is irreducible. Our goal is to show that $I(A)$ is prime. Consider $f, g \in K[\pi_1, \dots, \pi_n]$. Assume $fg \in I(A)$. Our goal is to show that $f \in I(A)$ or $g \in I(A)$. Since $fg \in I(A)$, hence $\{fg\} \subseteq I(A)$. Hence we deduce that $V(\{fg\}) \supseteq A$ by the equivalence proved in Definition 2.2.1, because (I^n, V^n) is a Galois connection. Hence we may argue as follows:

$$\begin{aligned} A &= A \cap V(fg) \\ &= A \cap (V(f) \cup V(g)) && \text{by Proposition 3.1.3} \\ &= (A \cap V(f)) \cup (A \cap V(g)) \end{aligned}$$

Since A is irreducible, we deduce that $A = A \cap V(f)$ or $A = A \cap V(g)$. Hence $A \subseteq V(f)$ or $A \subseteq V(g)$. Again exploiting Definition 2.2.1, we infer that $\{f\} \subseteq I(A)$ or $\{g\} \subseteq I(A)$, as desired. Hence $f \in I(A)$ or $g \in I(A)$, as desired.

(3) ⇒ (1). Assume $I(A)$ is a prime ideal. Assume toward a contradiction that A is not irreducible in the Zariski topology. Then there exist Zariski-closed proper subsets of A, call them X and Y, such that $A = X \cup Y$. Hence $I(A) = I(X \cup Y)$. Thus $I(A) = I(X) \cap I(Y)$. Thus $I(A) \supseteq I(X) \cdot I(Y)$. Since $I(A)$ is prime, we deduce $I(A) \supseteq I(X)$ or $I(A) \supseteq I(Y)$. Hence $A \subseteq X$ or $A \subseteq Y$. In the first case, we obtain a contradiction because, since X is a proper subset of A, hence $A \supsetneq X$. The second case is similar.

(2) ⇒ (3). If $A = V(\mathfrak{p})$ for some prime ideal $\mathfrak{p}$, then $I(A) = I(V(\mathfrak{p}))$. But by Proposition 3.1.7, this means $I(A) = \sqrt{\mathfrak{p}}$. But since every prime is a radical ideal, hence $I(A) = \mathfrak{p}$, which means that $I(A)$ is a prime ideal.

(3) $\Rightarrow$ (2). Assume $I(A)$ is a prime ideal. Since A is Zariski-closed, hence $A = V(\mathfrak{i})$ for ideal $\mathfrak{i}$. Hence $I(A) = I(V(\mathfrak{i}))$. Thus $I(A) = \sqrt{i}$. Hence $\sqrt{i}$ is prime. But this implies that $\mathfrak{i}$ is prime. Hence $A = V(\mathfrak{p})$ for some prime ideal $\mathfrak{p}$, namely $\mathfrak{p} := \mathfrak{i}$. □

Given a ring R, write Ideal(R) for the set of ideals of R. Given a topological space X, write $\mathcal{C}(X)$ for the closed subsets of X. To show that Zariski-closed sets can be decomposed as unions of irreducible Zariski-closed sets, the concepts Noetherian ring / Noetherian space are helpful.

Definition 3.2.3. Let P denote a poset. Then:

(a) P satisfies the *ascending chain condition* if and only if there are no order-preserving injections $\mathbb{N} \to P$.

(b) P satisfies the *descending chain condition* if and only if there are no order-reversing injections $\mathbb{N} \to P$.

Definition 3.2.4. The meaning of the word 'Noetherian' for both rings and topological spaces is defined below.

(a) A ring R is said to be *Noetherian* iff the poset Ideal(R) satisfies the ascending chain condition.

(b) A topological space X is said to be *Noetherian* iff the poset $\mathcal{C}(X)$ satisfies the descending chain condition.

Proposition 3.2.5 (Hilbert's Basis Theorem)**.** If a ring R is Noetherian, then so too is the polynomial ring $R[x]$.

Proof. This Theorem 2.7 in [27, p.37]. □

Proposition 3.2.6. For each natural number n, the space, the ring $K[\pi_1, \ldots, \pi_n]$ is Noetherian.

Proof. This can be proved by induction, using Proposition 3.2.5 for the inductive step. □

Proposition 3.2.7. For each natural number n, the space $\mathbb{A}^n_K$ is Noetherian.

Proof. Assume toward a contradiction that $\mathbb{A}^n_K$ were non-Noetherian. Then there exists an injective order-reversing mapping $j : \mathbb{N} \to \mathcal{C}(\mathbb{A}^n_K)$. Hence the injection $I \circ j : \mathbb{N} \to \text{Ideal}(K[\pi_1, \ldots, \pi_n])$ is order-preserving. Hence the ring $K[\pi_1, \ldots, \pi_n]$ is non-Noetherian. But this contradicts Proposition 3.2.6. □

Proposition 3.2.8. Let X denote a Noetherian topological space. Then there is a unique finite subset $F \subseteq \mathcal{C}(X)$ such that the following hold.

(a) Every $C \in F$ is irreducible.

(b) F is an antichain, meaning that if $A, B \in F$, then from $A \neq B$ we can deduce $A \nsubseteq B$ and $B \nsubseteq A$.

Proof. This is Proposition 2.31 in Milne [27, p.45]. □

For example, letting $X = V_2(\pi_1^2 - \pi_2^2)$ denote the space we looked at previously (with the subspace topology coming from $\mathbb{A}_K^2$), the F in the above theorem is

$$\{V(\pi_1 - \pi_2), V(\pi_1 + \pi_2)\}.$$

It's worth noting that although the above set appears to have two elements, in characteristic 2 it only has one element.

3.3 Affine varieties and regular mappings

Now that we know vanishing sets can be expressed as a finite union of certain basic building blocks, it makes sense to focus attention on this building blocks. This is what the notion of an affine varieties does for us. However, if we wish to follow Silverman's approach, a further issue needs to be addressed. Under our definitions, the set $V_{\mathbb{Q}}^2(\pi_2^2 - 2\pi_1^2)$ is irreducible. Geometrically, this is because, although it consists of two lines crossing at the origin, since those lines have irrational slope, the set cannot be decomposed as a union of the two lines, because those lines aren't the vanishing sets of any polynomials with coefficients in $\mathbb{Q}$. Hence from our point of view, this set is irreducible. However, Silverman defines things a little differently see [43, p.3], for example. Therefore, in order to ensure our approach better coincides with Silverman's, need to introduce the notion of *geometric irreducibility.*

Definition 3.3.1. Given a subset of affine space $S \subseteq \mathbb{A}_K^n$, we write $\overline{S}$ for exactly the same set, but regarded as a subset of the space $\mathbb{A}_{\overline{K}}^n$, which has the same elements but a different topology.

Definition 3.3.2. A set $S \subseteq \mathbb{A}_K^n$ is *geometrically irreducible* if and only if $\overline{S} \subseteq \mathbb{A}_{\overline{K}}^n$ is irreducible.

Remark. Our terminology above follows Poonen [36, p.7]. In general, a set $S \subseteq \mathbb{A}^n_K$ is sometimes said to be *geometrically whatever* if and only if $\overline{S}$ is *whatever*, and the above definition is the special case where the 'whatever' is irreducibility.

For example, the irreducible set $V^2_{\mathbb{Q}}(\pi_2^2 - 2\pi_1^2)$ considered earlier is not geometrically reducible, because:

$$\overline{V^2_{\mathbb{Q}}(\pi_2^2 - 2\pi_1^2)} = V^2_{\overline{\mathbb{Q}}}(\pi_2^2 - 2\pi_1^2) = V^2_{\overline{\mathbb{Q}}}(\pi_2 - \sqrt{2}\pi_1) \cup V^2_{\overline{\mathbb{Q}}}(\pi_2 + \sqrt{2}\pi_1).$$

An interesting question, not pursued in this thesis, is: 'To what extent does the theory of elliptic curves generalize to curves that are merely assumed irreducible, not geometrically irreducible?' We will not pursue this question much in this thesis, but it could easily be used as a springboard for further learning.

Definition 3.3.3. An *affine variety* X is a geometrically-irreducible subset of $\mathbb{A}^n_K$.

Every affine variety $X \subseteq \mathbb{A}^n_K$ can be regarded as a topological space with the subspace topology. Sometimes it's helpful to move to the algebraic closure; given $X \subseteq \mathbb{A}^n_K$, we'll write $\overline{X}$ for exactly the same set, but regarded as a subspace of $\mathbb{A}^n_{\overline{K}}$, and equipped with the subspace topology induced by this latter inclusion.

Affine varieties can be regarded up to equality, but it is usually more interesting to regard them up to isomorphism. To define a useful notion of morphism between affine varieties, a reasonable first attempt would be to try using continuous mappings. We quickly realize, however, that these are far too general. For example, the closed subsets of $\mathbb{A}^1_{\mathbb{C}}$ are precisely the finite subsets of $\mathbb{C}$ and $\mathbb{C}$ itself. But this means that for every function $f : \mathbb{C} \to \mathbb{C}$, if every $z \in \mathbb{C}$ has the property that $f^{-1}(z)$ is finite, then f is continuous when viewed as a mapping $\mathbb{A}^1_{\mathbb{C}} \to \mathbb{A}^1_{\mathbb{C}}$. This is far too general to be useful. Instead, we define morphisms of affine varieties as lists of formal polynomials.

Definition 3.3.4. If $A \subseteq \overline{K}^a$ and $B \subseteq \overline{K}^b$ are affine varieties, then a *regular mapping* $f : A \to B$ is a sequence $(f_1, \dots, f_b)$ of elements of $K[\pi_1, \dots, \pi_a]$ such that for all $a \in A$ we have $(f_1(a), \dots, f_b(a)) \in B$.

To reduce the notational overhead, we'll identify f and (f_1) whenever $b = 1$ in the above definition. For an example of a regular mapping, let $A = \{x, y \in \mathbb{C}^2 : x^2 + y^2 = 1\}$. Then we get a regular mapping (f_1, f_2) from A to itself by defining $f_1(x, y) = x^2 - y^2$ and $f_2(x, y) = 2xy$. However, the above definition is somehow incomplete, because it doesn't specify what it means for two regular mappings to be equal. Let's rectify that:

Definition 3.3.5. Two regular mappings

$$f, g : A \to B$$

are equal if and only if for all $a \in A$, we have $f(a) = g(a)$. We write $\mathrm{Hom}(A, B)$ for the set of regular mappings from A to B modulo the aforementioned equality relation.

Returning to the previous example, we see that the regular mapping (f^1, f_2) from A to itself given by $f^1(x, y) = 2x^2 - 1$ is equal to the mapping (f_1, f_2) described above. The important thing, then, is that two regular functions $f_1, f^1 \in \mathrm{Hom}(A, \mathbb{A}^1_K)$ are equal if and only if the function $f^1 - f_1 \in \mathrm{Hom}(A, \mathbb{A}^1_K)$ equals the zero function. Equality of regular functions $A \to B$ can then be reduced to an entrywise application of this observation. This motivates the following definition:

Definition 3.3.6. Given affine variety A, the coordinate ring $K[A]$ is defined as follows:

$$K[A] := \mathrm{Hom}(A, \mathbb{A}^1_K).$$

Remark. We'll write $\overline{K}[A]$ as shorthand for $\overline{K}[\overline{A}]$, since the meaning is clear. Explicitly:

$$\overline{K}[A] := \mathrm{Hom}(A, \mathbb{A}^1_{\overline{K}}).$$

Proposition 3.3.7. If $A \subseteq \mathbb{A}^n_K$ is an affine variety, then

$$K[A] \cong \frac{K[\pi_1, \ldots, \pi_n]}{I(A)}.$$

Proof. Define a surjective homomorphism as follows:

$$\begin{aligned} K[\pi_1, \ldots, \pi_n] &\xrightarrow{\rho_A} K[A] \\ f &\longmapsto f. \end{aligned}$$

Our goal is to use the definition of equality of regular mappings an affine variety A show that the kernel of ρ is precisely $I(A)$. Consider fixed but arbitrary $f \in K[\pi_1, \ldots, \pi_n]$. Our goal is to show that

$$f \in \ker(\rho_A) \iff f \in I(A).$$

We have:

$$\begin{aligned} \text{LHS} &\iff f \in \ker(\rho_A) \\ &\iff \rho_A(f) = 0 \\ &\iff V(f) \supseteq A \\ &\iff f \in I(A) && \text{by Definition 2.2.1} \\ &\iff \text{RHS}. \end{aligned}$$

We deduce that $\ker(\rho_A) = I(A)$, as desired. □

As a special case, we obtain:

Corollary 3.3.8. For each natural number n, we have:

$$K[\mathbb{A}_K^n] = K[\pi_1, \ldots, \pi_n].$$

It's also worth noting that we have a fairly explicit description of $\mathrm{Hom}(A, \mathbb{A}_K^b)$ in terms of coordinate rings. In particular:

Proposition 3.3.9. For each natural number b,

$$\mathrm{Hom}(A, \mathbb{A}_K^b) = K[A]^b.$$

Proof. Both $\mathrm{Hom}(A, \mathbb{A}_K^b) = \mathrm{Hom}(A, \mathbb{A}_K^1)^b$ and $\mathrm{Hom}(A, \mathbb{A}_K^1) = K[A]$ are true by definition. □

3.3.1 Localization and multiplicity

Sometimes polynomials aren't general enough, and we want to allow functions to 'blow up' everywhere except the point $P \in A$ that we're considering. This can be achieved by localizing at the relevant prime ideal.

Definition 3.3.10. Let $P \in A$ denote a point in an affine variety. Then the local ring at P, denoted $K[A]_P$, is the localization of the ring $K[A]$ at the prime ideal $I(P)$ of all polynomials that vanish at P.

In other words, the denominators are required to *not vanish* at P. In symbols:

$$K[A]_P = \{f/g : f \in K[A], g \in K[A], g(P) \neq 0\}$$

We'll also write $K(A)$ for the field of fractions of $K[A]$, which can be defined as the localization of $K[A]$ at the prime ideal $\{0_{K[A]}\}$. Thus:

$$K[A]_P = \{f/g : f \in K[A], g \in K[A], g \neq 0\}$$

Note that, since A is an affine variety, thus $I(A)$ is prime ideal, and hence $K[A]$ is an integral domain, and thus the function

$$\begin{aligned} K[A] &\to K(A) \\ f &\mapsto \frac{f}{1} \end{aligned}$$

is injective. We'll tend to regard $K[A]$ and $K[A]_P$ as subsets of $K(A)$ for this reason.

Definition 3.3.11. Let $P \in A$ denote a point in an affine variety. Then M_P denotes the set of all polynomials on A that vanish at P. Explicitly:

$$M_P = \{f \in K[A] : f(P) = 0\}.$$

Given a point $P \in A \subseteq \mathbb{A}^n_K$, we'll write $\overline{P} \in \overline{A}$ for the same point, but regarded as a subset of $\overline{A} \subseteq \mathbb{A}^n_{\overline{K}}$. This means, in particular, that $M_{\overline{P}} = \{f \in \overline{K}[A] : f(P) = 0\}$.

Proposition 3.3.12. Let $P \in A$ denote a point in an affine variety. Then $K[A]_P$ is a Noetherian local ring with maximal ideal M_P.

Proof. See [4] for, example. □

This allows us to apply Krull's intersection theorem to understanding the structure of $K[A]_P$.

Proposition 3.3.13 (Krull's Intersection Theorem)**.** Let R be a noetherian local ring with maximal ideal M. Then for all non-zero $f \in R$, there exists an $n \in \mathbb{N}$ such that $f \notin M^n$.

Proof. This is Theorem 1.8 in Milne [27, p.15]. □

This motivates the following definition:

Definition 3.3.14. Let C be an affine variety and suppose $P \in C$ is a smooth point. Then we obtain a function ord_P defined as follows:

$$\begin{aligned} \overline{K}[C]_P &\xrightarrow{\mathrm{ord}_P} \mathbb{N} \cup \{\infty\} \\ f &\longmapsto \sup\{d \in \mathbb{N} : f \in M_P^d\}. \end{aligned}$$

We extend this to negative values as follows, whenever such a function exists:

$$\begin{aligned} \overline{K}(C) &\xrightarrow{\mathrm{ord}_P} \mathbb{Z} \cup \{\infty\} \\ f/g &\longmapsto \mathrm{ord}_P(f) - \mathrm{ord}_P(g). \end{aligned}$$

Note that by the Krull Intersection Theorem, the condition $\mathrm{ord}_P(f) = \infty$ is equivalent to $f = 0$ for all $f \in \overline{K}[C]_P$. This means that, for non-zero f, we get an integral value for $\mathrm{ord}_P(f)$.

3.4 Dimension of an affine variety

Finding a purely topological definition of the dimension of a space that gives the right answer with respect to the standard topology is, in general, a non-trivial task. The interested reader is directed to Alan Pear's authoritative work on the subject [34]. However when the spaces under consideration have the Zariski topology, the notion of Krull dimension provides a straightforward topological definition.

Definition 3.4.1. Let P denote a poset. Then the *length* of P is the supremum of all $n \in \mathbb{N} \cup \{\infty\}$ such that there exists an injective monotone function $\eta : \{0, \ldots, n\} \to P$.

Definition 3.4.2. We make the following definitions pertaining to Krull dimension.

(a) Let X denote a Noetherian topological space. Then $\mathcal{I}(X)$ denotes the poset of all irreducible closed subsets of X. The *Krull dimension* of the space X is the length of the poset $\mathcal{I}(X)$.

(b) Let R denote a ring. Then $\mathrm{Spec}(R)$ denotes the poset of all prime ideals of R. The *Krull dimension* of R is the length of the poset $\mathrm{Spec}(R)$.

For example, the Krull dimension of the space $\mathbb{A}^2_K$ is easily seen to be at least 2, as seen by taking η_0 equal to a point, η_1 equal to a line containing that point, and η_2 equal to $\mathbb{A}^2_K$. Similarly, the Krull dimension of the ring $K[\pi_1, \pi_2]$ is easily seen to be at least 2, by taking η_0 equal to (0), η_1 to equal (π_1), and η_2 to equal (π_1, π_2). Of course, by the correspondence between prime ideals in $K[\pi_1, \ldots, \pi_n]$ and irreducible subsets of $\mathbb{A}^n_K$, both viewpoints will always give the same answer. What remains to show is that the Krull dimension of $\mathbb{A}^n_K$ is no greater than n. This follows from the strong form of Krull's principal ideal theorem; the interested reader is directed to Mel Hochester's notes on the subject [21]. We'll take a different approach however, and exploit the connections between Krull dimension and transcendence degree.

Proposition 3.4.3. A K-ring R is said to be of *finite type* if and only if there exists a finite set $\{r_1, \ldots, r_n\} \subseteq R$ such that for all $r \in R$, there exists a polynomial $f \in K[\pi_1, \ldots, \pi_n]$ such that $r = f(r_1, \ldots, r_n)$.

Proposition 3.4.4. Let R denote a K-ring of finite type that is also an integral domain. Then the transcendence degree of the field of fractions of R over K equals the Krull dimension of K.

Proof. See Theorem 5.6.7 in Chapter 5 of Robert Ash's Commutative Algebra text [2, p.15]. □

We're now in a position to prove:

Corollary 3.4.5. The Krull dimension of $\mathbb{A}^n_K$ is precisely n.

Proof. This is equivalent to showing that the Krull dimension of the ring $K[\pi_1, \ldots, \pi_n]$ is exactly n. Hence by Proposition 3.4.4, it is enough to show that the transcendence degree of $K(\pi_1, \ldots, \pi_n)$ over K is exactly n. But this is Corollary 1.5.4. □

Having thus (at least partially) justified Krull dimension as a sensible notion of dimension for the purposes of algebraic geometry, let us hereafter simply refer to this as the *dimension* of an affine variety.

Definition 3.4.6. Given an affine variety X, write $\dim(X)$ for the dimension of X.

3.5 Smoothness

Following Silverman [43, p.4], we define smoothness via partial derivatives as follows:

Definition 3.5.1. Given an affine variety $X \subseteq \mathbb{A}^n_K$, we say that X is *smooth* at $P \in X$ if and only if there exists a sequence $f = (f_1, \ldots, f_r)$ of elements of $\overline{K}[\pi_1, \ldots, \pi_n]$ such that firstly, the ideal generated by $\{f_1, \ldots, f_r\}$ equals $I_{\overline{K}}(X)$, and secondly, the matrix of partial derivatives of f evaluated at P has rank $n - \dim(X)$.

There are many equivalent characterizations of this property. For example:

Definition 3.5.2. Let X be a variety. A point $P \in X$ is smooth if and only if $\dim_{\overline{K}}(M_{\overline{P}}/M^2_{\overline{P}}) = \dim(X)$.

Proof. This is Proposition 1.7 in Silverman [43, p.5]. □

Definition 3.5.3. An affine variety A is said to be smooth if and only if every point $P \in A$ is smooth.

3.6 Affine curves

Definition 3.6.1. An *affine curve* is a 1-dimensional affine variety.

For example, the hyperbola $V_K^2(\pi_1^2 - \pi_2^2 - 1)$ is an affine curve. So too is the curve $V_K^2(\pi_1^2 - \pi_2^3)$ displayed in Figure 4.1; notice this curve has a cusp as the origin, and is thus an example of a non-smooth curve (the notion of a cusp can be made precise, though we will not do so here). Another example is the line

$$V^3(\pi_1, \pi_2) = V^3(\pi_1) \cap V^3(\pi_2) = \{(x, y, z) \in K^3 : x = 0 \wedge y = 0\} = \{(0, 0, z) : z \in K\},$$

which indicates that affine curves needn't be a subset of affine 2-space.

Definition 3.6.2. An *affine plane curve* (or *planar affine curve*) is an affine curve that's a subset of $\mathbb{A}_K^2$.

We've already seen a few examples; a simple non-example is the set $V_K^2(\pi_1^2 - \pi_2^2)$, which, being a union of two distinct lines, is not affine variety. The set $V_{\mathbb{Q}}^2(\pi^2 - 2\pi^2)$ is also a non-example; though irreducible, it decomposes into two lines over $\overline{\mathbb{Q}}$.

3.7 Projective space, projective varieties and projective curves

Many constructions in geometry become simpler when moved from affine space to projective space, defined as follows:

Definition 3.7.1. For each natural number n, projective n-space, denoted $\mathbb{P}_K^n$, is defined as the following quotient

$$\mathbb{P}^n(K) = \frac{\{P \in \overline{K}^{n+1} : P \neq 0\}}{(P, Q) \mapsto \exists k \in \overline{K}^* : kP = Q}.$$

Projective space is a fundamental object of study in algebraic geometry, but a proper development is out of the scope of this thesis due to time constraints. The interested reader is directed to *Elliptic Tales* by Avner Ash and Robert Gross [1], which includes, among other things, a remarkably pedagogical account of the basics of projective space.

We proceed quickly but informally. A graded ideal of $K[\pi_0, \ldots, \pi_n]$ is an ideal generated by homogeneous polynomials. Given a graded ideal $\mathfrak{a}$ of $K[\pi_0, \ldots, \pi_n]$, we write $V_K^n(\mathfrak{a})$ for the projective vanishing set of $\mathfrak{a}$. This allows us to equip projective space with a topology that is also called the Zariski topology. The geometrically irreducible sets in this topology are called *projective varieties*, and those with Krull dimension 1 are called projective curves. A *planar projective curve* is a projective curve that is also a subset of $\mathbb{P}^2$.

Projective n-space can be covered by charts that are isomorphic to affine n-space. In this way, we can apply most of what was developed for affine varieties to the projective world. For example, if given $P \in X$ is a point of a projective variety, we say that P is smooth if and only if there exists a chart containing P such that P is smooth when regarded as an element of the corresponding affine variety. The reader is directed to Silverman [43, p.9] for a definition of the basic charts. We can define $\operatorname{ord}_f(P)$ in essentially the same way. We'll tend to distinguish one of these charts and regard it as an inclusion to ease the notation. In particular, we regard A_K^n as a subset of $\mathbb{P}_K^n$ via the function

$$\begin{aligned} \overline{K}^n &\to \mathbb{P}^n \\ (P_1, \dots, P_n) &\mapsto [P_1 : \cdots : P_n : 1]. \end{aligned}$$

Given a projective variety X, we follow Silverman [43, p.10] in defining the field of rational functions on X as $K(X) := K(X \cap \mathbb{A}_K^n)$. The following result will be helpful:

Proposition 3.7.2. Let $f : X \setminus \{P\} \to Y$ denote a regular function, where X is a projective curve and Y is a projective variety. Then there exists a unique regular function $\tilde{f} : X \to Y$ such that $\tilde{f} \restriction_{X \setminus \{P\}} = f$.

Proof. See Proposition 6.8 in Hartshorne [20, p.137] □

3.8 Regular functions between projective curves

Proposition 3.8.1. Let $f : C_1 \to C_2$ denote a regular function between projective curves. Then either f is constant, or f is surjective.

Proof. This is Theorem 2.3 in Silverman [43, p.20]. □

This allows us to make the following definition.

Definition 3.8.2. Let $\varphi : C_1 \to C_2$ denote a non-constant regular function between projective curves. Then we obtain a K-algebra homomorphism $K(\varphi) : K(Y) \to K(X)$

by defining

$$K(\varphi)(f) = f \circ \varphi$$

for all $f \in K(Y \cap \overline{K}^n)$.

Definition 3.8.3. Let $\varphi : X \to Y$ denote a regular function between projective curves. Then φ is *separable* if and only if $K(\varphi)$ is separable.

Proposition 3.8.4. Given two morphisms of projective curves

$$C_1 \xrightarrow{\varphi} C_2 \xrightarrow{\psi} C_3,$$

if φ and ψ are non-constant, then so too is their composite $\psi \circ \varphi$.

Proof. Then they're both surjective by Proposition 3.8.1, and hence $\psi \circ \varphi$ is surjective. Thus since C_3 is a curve, it strictly more than 1 point, and we deduce that $\psi \circ \varphi$ is non-constant. □

Proposition 3.8.5. Let $\varphi : C_1 \to C_2$ denote a morphism of projective curves. If φ is non-constant, then the corresponding morphism of fields $K(\varphi) : K(C_2) \to K(C_1)$ has finite degree.

Proof. This is Theorem 2.4 in Silverman [43, p.20]. □

Definition 3.8.6. Let $\varphi : C_1 \to C_2$ denote a regular function between projective curves. If φ is non-constant, we define the degree, the separable degree and the inseparable degree of φ as follows:

If φ is non-constant:

$$\begin{aligned} \deg(\varphi) &:= \deg(K(\varphi)) \\ \deg_s(\varphi) &:= \deg_s(K(\varphi)) \\ \deg_i(\varphi) &:= \deg_i(K(\varphi)) \end{aligned}$$

If φ is constant:

$$\begin{aligned} \deg(\varphi) &:= 0 \\ \deg_s(\varphi) &:= 0 \\ \deg_i(\varphi) &:= 0 \end{aligned}$$

Remark. By Proposition 3.8.5, these are always natural numbers.

The separable degree of morphism $\varphi : C_1 \to C_2$ of projective curves is especially interesting, due to its close connection with the number of points in preimage $\varphi^{-1}(Q)$ for most $Q \in C_2$. In particular:

Proposition 3.8.7. Let $\varphi : C_1 \to C_2$ denote a nonconstant morphism of projective curves. Then for all but finitely many $Q \in C_2$, we have $\#\varphi^{-1}(Q) = \deg_s(\varphi)$.

Proof. This is Part (b) of Proposition 2.6 in Silverman [43, p.24]. □

Proposition 3.8.8. Given two morphisms of projective curves

$$G \xrightarrow{\varphi} H \xrightarrow{\psi} I,$$

the degree function behaves as follows:

$$\deg(\psi \circ \varphi) = \deg(\varphi) \cdot \deg(\psi).$$

Proof. There are two cases. If at least one of φ or ψ is constant, then so too is $\psi \circ \varphi$, and hence both sides of the formula are 0 and the equality holds. So assume neither φ nor ψ is constant. Then neither is $\psi \circ \varphi$ by Proposition 3.8.4. Hence we may argue as follows:

$$\begin{aligned}
\text{LHS} &= \deg(\psi \circ \varphi) \\
&= \deg(K(\psi \circ \varphi)) && \text{because } \psi \circ \varphi \text{ is non-constant} \\
&= \deg(K(\varphi) \circ K(\psi)) \\
&= \deg(K(\varphi)) \cdot \deg(K(\psi)) && \text{by Proposition 1.4.2} \\
&= \deg(\varphi) \cdot \deg(\psi) && \text{because } \varphi \text{ and } \psi \text{ are non-constant} \\
&= \text{RHS}
\end{aligned}$$

□

3.9 Regular differential forms and genus

The material in this section is taken from Andries E. Brouwer's notes on the subject [8].

Definition 3.9.1. Let C denote a smooth projective curve. Then a *candidate for* $\Omega[C]$ is a pair (X, d) such that X is a $\overline{K}[C]$-module and $d : \overline{K}[C] \to X$ is a function satisfying the following axioms:

1. $d(f + g) = df + dg$

2. $d(fg) = df \cdot g + f \cdot df$

3. $da = 0$ if $a \in \overline{K}$

Candidates for $\Omega[C]$ form a category by defining that a morphism

$$f : (X, d_X) \to (Y, d_Y)$$

is a $\overline{K}[C]$-linear map $f : X \to Y$ satisfying $d_Y \circ f = d_X$.

Remark. The above axioms imply that $d : \overline{K}[C] \to X$ is $\overline{K}$-linear. To see this, note that since we already know that d is additive, it suffices to show that $d(af) = adf$ for all $a \in \overline{K}$. We compute $d(af) = da \cdot f + a \cdot df = 0 \cdot f + a \cdot df = adf$, as desired.

Definition 3.9.2. The notation $\Omega[C]$ is used for the initial object of the resulting category. Elements of $\Omega[C]$ are called *regular differential forms* on C, the vector space $\Omega[C]$ itself is referred to as the module of regular differential forms on C.

Definition 3.9.3. The *genus* of a smooth projective curve C is defined as

$$g_c := \dim_{\overline{K}} \Omega[C].$$

In this case where $\overline{K} = \mathbb{C}$, this agrees with the usual notion of genus used in low-dimensional topology.

3.10 Divisors and Picard groups

The idea behind the Picard group of an algebraic variety is as follows. Given an algebraic variety V, we may speak of the vector bundles on V, which can be defined as locally-free coherent sheaves on V. The one-dimensional vector bundles are called the *line bundles on* V, and can be characterized among the vector bundles as follows. We firstly observe that the category of vector bundles over V is naturally equipped with a notion of tensor product which makes it into a monoidal category. The unit in the aforementioned monoidal category is the trivial line bundle, and it can be shown that an object of this category has an inverse with respect to the tensor product if and only if it is a line bundle. So in some sense, ‘line bundle’ means ‘invertible element,’ and the set of isomorphism classes of line bundles over V becomes a group, with the law of composition taken as the (decategorified) tensor product. This group is called the *Picard group* of V, and denoted $\mathrm{Pic}(V)$.

However, although the above viewpoint is conceptually clean, it is also unnecessarily sophisticated for our purposes here. In particular, it can be shown that for smooth algebraic curves, the Picard group is naturally isomorphic to the set of all formal $\mathbb{Z}$-linear combinations of points on the curve, modulo a particular equivalence relations. We’ll follow Silverman [43, p.28] in taking this approach, as it is simpler to define, easier to compute with, and for smooth curves, yields exactly the same group. Our basic definitions are therefore as follows.

Definition 3.10.1. Let C denote a smooth projective curve. Then:

- A divisor of C is a formal $\mathbb{Z}$-linear combination of elements of C. These form an abelian group denoted $\mathbb{Z}\langle C\rangle$.

- Given $f \in K(C)^*$, we define $\operatorname{div}(f) \in \mathbb{Z}\langle C\rangle$ as follows:

$$\operatorname{div}(f) = \sum_{c\in C} \operatorname{ord}_c(f)(c).$$

- The notion *principal divisor on* C is the image of the function $\operatorname{div} : K(C)^* \to \mathbb{Z}\langle C\rangle$ defined above. In other words, a divisor $D \in \mathbb{Z}\langle C\rangle$ is said to be *principal* if and only if there exists $f \in K(C)^*$ such that $D = \operatorname{div}(f)$.

- The Picard group of C, denoted $\operatorname{Pic}(C)$ is the quotient of $\mathbb{Z}\langle C\rangle$ by the subgroup of principle divisors. The elements of $\operatorname{Pic}(C)$ can be thought of as isomorphism classes of line bundles on C, or equivalently, as divisors of C modulo principal divisors.

Definition 3.10.2. Given a projective curve C, each divisor D of C gives rise to a $\overline{K}$-module called Riemann-Roch space associated to D, defined as follows:

$$\mathcal{L}(D) = \{f \in \overline{K}(C)^* : \operatorname{div}(f) + D \geq 0\} \cup \{0\}.$$

Proposition 3.10.3. Given a projective curve C and a divisor D, the space $\mathcal{L}(D)$ is finite-dimensional $\overline{K}$-vector space.

Proposition 3.10.4. See Proposition 5.2 in Silverman [43, p.34].

A deep principle called the Riemann-Roch theorem controls the dimension of Riemann-Roch space. For our purposes here, we'll just focus on the following corollary of the full theorem. Define that the degree of a divisor D, denoted $\deg(D)$, is the sum of all the coefficients of D. In this notation, we have:

Proposition 3.10.5 (Weak form of Riemann-Roch theorem)**.** If C is a projective curve and D is a divisor satisfying $\deg(D) > 2g_C - 2$, then

$$\dim_{\overline{K}}\mathcal{L}(D) = \deg D - g_C + 1.$$

Proof. See Corollary 5.5 in Silverman [43, p.35]. □

Corollary 3.10.6. If C is a projective curve of genus 1 and D is a divisor satisfying $\deg(D) \geq 1$, then

$$\dim_{\overline{K}} \mathcal{L}(D) = \deg D.$$

Proposition 3.10.7. Let C denote a smooth projective curve. Then every principal divisor D on C has degree exactly 0.

Proof. Let D denote a principal divisor. Our goal is to show that $\deg(D) = 0$. Let $f \in K(C)$ satisfy $\operatorname{div}(f) = D$, so that our goal becomes showing that $\deg(\operatorname{div}(f)) = 0$. Since $f \in K(C)$, there is a natural number n together with regular functions $g, h \in K[C]_n$ such that $f = \frac{g}{h}$ Thus:

$$\operatorname{div}(f) = \operatorname{div}\left(\frac{g}{h}\right) = \operatorname{div}(g) - \operatorname{div}(h).$$

It follows that

$$\deg(\operatorname{div}(f)) = \deg(\operatorname{div}(g)) - \deg(\operatorname{div}(f)) = n - n = 0. \qquad \square$$

Corollary 3.10.8. If C is a smooth projective curve, then the degree function $\deg : \mathbb{Z}\langle C \rangle \to \mathbb{Z}$ descends to a function

$$\deg : \operatorname{Pic}(C) \to \mathbb{Z}.$$

Definition 3.10.9. Let C denote a smooth projective curve. Then the Picard-0 group of C is the group

$$\operatorname{Pic}^0(C) := \{D \in \operatorname{Pic}(C) : \deg(D) = 0\}.$$

Chapter 4

Elliptic Curves

The material in this chapter mainly comes from Silverman's book on elliptic curves [43].

Definition 4.0.1. An elliptic curve over a field K is a smooth projective curve of genus 1 with a distinguished K-rational point.

The purpose of including a distinguished point in the definition is that it will be used to endow each elliptic curve with a group structure via the Picard group.

Proposition 4.0.2. Suppose C is a smooth projective curve of genus 1. Then from $[(P)] = [(Q)]$ we may deduce $P = Q$.

Proof. Since $[(P)] = [(Q)]$, thus $[(P) - (Q)] = 0$. So there exists $f \in K(C)^*$ such that $\operatorname{div}(f) = (P) - (Q)$. Since $f \in \mathcal{L}((Q))$, hence f is a constant function by Corollary 3.10.6. It follows that $\operatorname{div}(f) = 0$. Thus $0 = (P) - (Q)$. So $(P) = (Q)$, from which we deduce $P = Q$. □

Now given a pointed projective curve (C, O), we get a corresponding map

$$\begin{aligned} C &\xrightarrow{\kappa_C} \operatorname{Pic}^0(C) \\ P &\longmapsto [(P) - (O)] \end{aligned}$$

where $[\square] : \mathbb{Z}\langle C\rangle \to \operatorname{Pic}(C)$ is the relevant quotient map.

Proposition 4.0.3. Let $E := (E, O)$ denote an elliptic curve. Then the function κ_E defined above is bijective.

Proof. (Surjectivity.) Consider $D \in \mathbb{Z}\langle E\rangle$ such that $\deg(D) = 0$. Our goal is to find $P \in E$ such that $[\kappa_E(P)] = [D]$. By definition, this equation reads $[(P) - (O)] = [D]$, which reads

$$[(P)] = [D + (O)].$$

Since $\deg(D + (O)) = 1$ thus by Proposition 3.10.6, we deduce that

$$\dim(\mathcal{L}(D + (O))) = \deg(D + (O)) = \deg(D) + \deg(O) = 0 + 1 = 1.$$

So let $f \in \mathcal{L}(D+(O))$ have the property that $\{f\}$ is a basis. Then from the definition of Riemann-Roch space, we have

$$\operatorname{div}(f) + D + (O) \geq 0.$$

Now compute

$$\deg(\operatorname{div}(f) + D + (O)) = 0 + 0 + 1 = 1.$$

Hence there exists $P \in E$ such that

$$\operatorname{div}(f) + D + (O) = (P).$$

Hence $[(P)] = [D + (O)]$, as desired.

(Injectivity.) Consider points $P, Q \in E$. Assume $\kappa_E(P) = \kappa_E(Q)$. Our goal is to show that $P = Q$. Since $\kappa_E(P) = \kappa_E(Q)$, thus $[(P) - (O)] = [(Q) - (O)]$. Hence $[P] = [Q]$. But by Proposition 4.0.2 this implies $P = Q$, as desired. □

The above result allows us to put a group structure on each elliptic curve by transporting the group structure from $\operatorname{Pic}^0(E)$ to E. Explicitly:

Definition 4.0.4. Let (E, O) denote an elliptic curve. Then we define functions on E as follows:

$$\begin{aligned} E \times E &\xrightarrow{\square \oplus \square} E & \qquad E &\xrightarrow{-\square} E \\ (P, Q) &\longmapsto \kappa_E^{-1}(\kappa_E(P) + \kappa_E(Q)) & \qquad P &\longmapsto \kappa_E^{-1}(-\kappa_E(P)) \end{aligned}$$

Proposition 4.0.5. If (E, O) is an elliptic curve, then $(E, \oplus, O, -\square)$ is an abelian group.

Proof. This follows from some elementary computations. □

It's not immediately obvious, but the above functions are in fact regular; this means, in particular, that elliptic curves are abelian varieties:

Definition 4.0.6. An *abelian variety* is an abelian group object in the category of projective varieties; that is, it's a projective variety X together with a point $O \in X$ and regular mappings

$$X \times X \xrightarrow{\square \oplus \square} X \qquad\qquad X \xrightarrow{-\square} X$$

satisfying the equational abelian group axioms, namely:

$$(P + Q) + R = P + (Q + R)$$

$$O + P = P,\ P + O = P$$

$$P + -P = O,\ -P + P = O$$

$$P + Q = Q + P$$

Remark. The reader may prefer to omit some of the above axioms to prevent redundancy.

To see that the functions defined above make each elliptic curve into an abelian variety, there are a couple of ways to proceed. In order to avoid unnecessary abstraction, in this thesis we'll take the most direct approach. In particular, our strategy is to show that every elliptic curve can be put into a particular normal form called Weierstrass form. The group law on a Weierstrass curve can be written down directly and is immediately seen to be given by regular functions. In this way, we'll deduce that every elliptic curve is an abelian variety.

4.1 Isogenies

Elliptic curves form a category in which the morphisms are referred to as *isogenies.*

Definition 4.1.1. Let (E_0, O_0) and (E_1, O_1) denote elliptic curve. Then an *isogeny*

$$\varphi : (E_0, O_0) \to (E_1, O_1)$$

is a regular mapping $\varphi : E_0 \to E_1$ such that $\varphi(O_0) = O_1$.

Being an isogeny may not look like a very restrictive condition upon first glance, but many results follow from the definition. We collect the most important such results here below.

Proposition 4.1.2. Let $\varphi : (E_0, O_0) \to (E_1, O_1)$ denote an isogeny. Then φ is a group homomorphism $(E_0, \oplus) \to (E_1, \oplus)$.

Proof. I claim firstly that the following diagram commutes

$$\begin{tikzcd}
E_0 \arrow[r, "\kappa_{E_0}"] \arrow[d, "\varphi"'] & \mathrm{Pic}^0(E_0) \arrow[d, "\mathrm{Pic}^0(\varphi)"] \\
E_1 \arrow[r, "\kappa_{E_1}"] & \mathrm{Pic}^0(E_1).
\end{tikzcd}$$

Our goal is to show that $\mathrm{Pic}^0(\varphi) \circ \kappa_{E_0} = \kappa_{E_1} \circ \varphi$. Consider $P \in E_0$. Then the left-hand-side can be reduced as follows.

$$\begin{aligned}
\text{LHS} &= (\mathrm{Pic}^0(\varphi) \circ \kappa_{E_0})(P) \\
&= \mathrm{Pic}^0(\varphi)(\kappa_{E_0}(P)) \\
&= \mathrm{Pic}^0(\varphi)((P) - (O)) \\
&= (\varphi(P)) - (\varphi(O)) \\
&= (\varphi(P)) - (O)
\end{aligned}$$

The right-hand-side is similarly reduced.

$$\begin{aligned}
\text{RHS} &= (\kappa_{E_1} \circ \varphi)(P) \\
&= \kappa_{E_1}(\varphi(P)) \\
&= (\varphi(P)) - (O)
\end{aligned}$$

Since LHS = RHS, hence the above diagram commutes. We deduce that

$$\varphi = \kappa_{E_1}^{-1} \circ \mathrm{Pic}^0(\varphi) \circ \kappa_{E_0}$$

Hence φ, being a composite of group homomorphisms, is itself a group homomorphism. □

Definition 4.1.3. If E_0 and E_1 are elliptic curves, the zero isogeny is the function

$$\begin{tikzcd} E_0 \arrow[r, "0em"] & E_1 \\ P \arrow[r, mapsto] & O. \end{tikzcd}$$

Proposition 4.1.4. If E_0 and E_1 are elliptic curves, every non-zero isogeny from E_0 to E_1 is surjective.

Proof. This follows from Proposition 3.8.1. □

The above observation suggests that there is no interesting notion of sub-elliptic curve. In particular, suppose hypothetically that we distinguished certain subsets of E as sub-elliptic curves. Suppose our definition designed so that each sub-elliptic curve becomes equipped with the structure of an elliptic curve (inherited from E) in a deterministic way, such that the inclusion to E becomes an isogeny. Then, in this case, there can be no sub-elliptic curves other than E and $\{O\}$. For suppose S is a sub-elliptic curve of E with a point distinct from O. Then the inclusion $S \hookrightarrow E$ is a non-constant isogeny, and hence surjective, and hence $S = E$. This argument suggests that no attempts at defining the phrase 'sub-elliptic curve' will succeed at producing a useful and non-trivial notion.

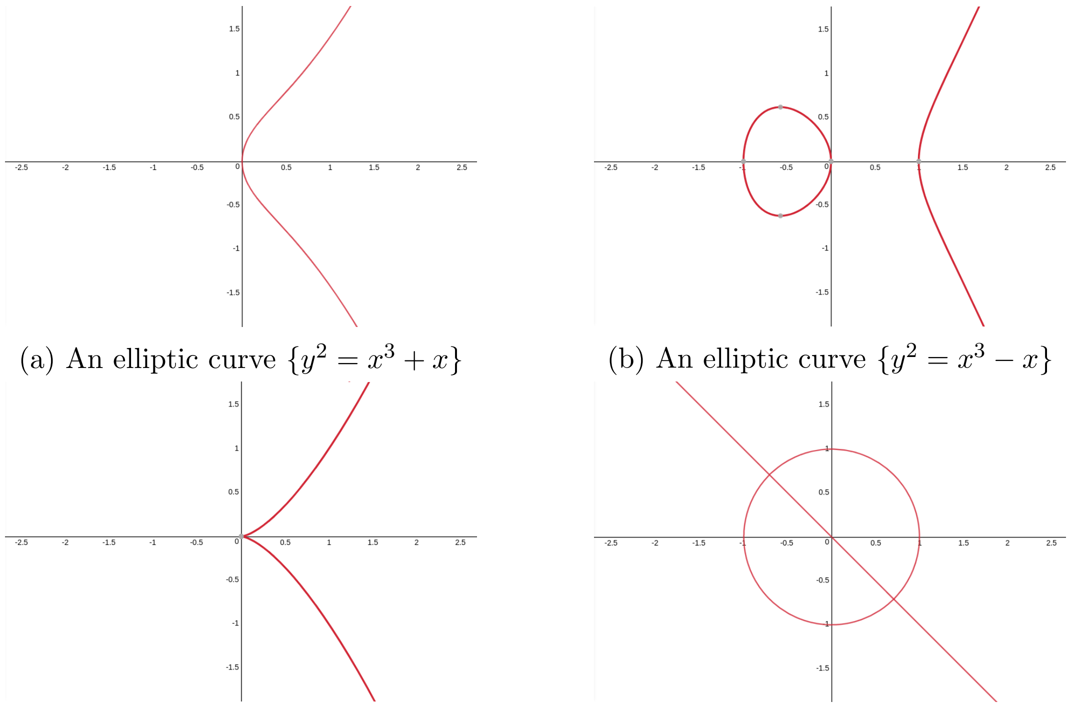
4.2 Planar elliptic curves

In general, we can get elliptic curves in the following way. Recall that for a projective plane curve, the genus-degree formula relates the degree of the defining polynomial with the curve's genus. In particular, the formula says that given a projective plane curve C, if d is the degree of C and g is its (arithmetical) genus, then

$$g = \frac{1}{2}(d-1)(d-2).$$

Note that our definition of curve includes an assumption of irreducibility, and that the above formula requires this assumption.

Note also that, in general, the g must be regarded as the arithmetical genus. However, for smooth curves this coincides with the geometric genus, and hence there is no ambiguity in the meaning of the word 'genus.' What this means is that for smooth projective plane curves of degree d, being of genus 1 is equivalent to $2 = (d-1)(d-2)$. Since $d := 2$ is too small for this to be true, and $d := 4$ is too large, hence the only



(a) An elliptic curve $\{y^2 = x^3 + x\}$

(b) An elliptic curve $\{y^2 = x^3 - x\}$

(c) A non-smooth cubic planar curve $\{y^2 = x^3\}$

(d) The set $\{(x + y)(x^2 + y^2 - 1) = 0\}$

Figure 4.1: The affine portion of some cubic algebraic subsets of $\mathbb{P}^2$

solution is $d := 3$. We deduce that if we're dealing with projective plane curves, the only elliptic curves are those of degree exactly 3.

Therefore, to get examples of elliptic curves, one approach is to choose a cubic polynomial in three variables satisfying certain requirements. That is, we look for a non-zero K-linear combination f of the monomials

$$x^3, x^2y, xy^2, y^3, x^2z, xyz, y^2z, xz^2, yz^2, z^3.$$

To obtain an elliptic curve, we need to make sure that f is irreducible, to avoid examples like (d) in Figure 4.1. We also need to make sure that the vanishing set of f is smooth, to avoid examples like (c) in Figure 4.1. Once these conditions are satisfied, the genus-degree formula takes care of the rest and ensures that the vanishing set of f becomes an elliptic curve upon choosing a basepoint.

In the special case where the elliptic curve (E, O) we're interested in is a cubic planar curve, the group law defined in the previous section can be described explicitly via a geometric construction involving lines and points of intersection.

Definition 4.2.1. Given a homogeneous cubic polynomial $f \in \overline{K}[x, y, z]$, if f is irreducible and $V(f)$ is smooth, then given $P, Q \in V(f)$, we obtain a corresponding line

$$P * Q \subseteq \mathbb{P}^2$$

in the following way.

- If $P \neq Q$, then $P * Q$ is the unique line that contains both P and Q.
- If $P = Q$, then $P * Q$ is the tangent line to $V(f)$ at P.

Definition 4.2.2. Given a homogeneous cubic polynomial $f \in \overline{K}[x, y, z]$, if f is irreducible and $V(f)$ is smooth, then given $P, Q \in V(f)$, the notation $P \star Q$ refers to the unique element of $V(f)$ satisfying

$$(P \star Q) = \operatorname{div}_E(P * Q) - (P) - (Q).$$

Remark. The rearranged from $\operatorname{div}_E(P * Q) = (P) + (Q) + (P \star Q)$ will be helpful in what follows.

Proposition 4.2.3. Let E denote a smooth cubic subset of $\mathbb{P}^2$ and assume $O \in E$ is a K-rational point. Then for all $P, Q \in E$, we have:

$$P \oplus Q = (P \star Q) \star O$$

Proof. Recall that each elliptic curve E is associated with a bijection $\kappa : E \to \operatorname{Pic}^0(E)$ defined by $\kappa(P) = (P) - (O)$. Recall also that $\oplus$ is defined by $P \oplus Q = \kappa^{-1}(\kappa(P) + \kappa(Q))$. By the definition of $\oplus$, we wish to show that

$$\kappa(P) + \kappa(Q) = \kappa((P \star Q) \star O).$$

That is, we're trying to show that

$$[(P) - (O)] + [(Q) - (O)] = [((P \star Q) \star O) - (O)].$$

Define α to to be the RHS of the above expression minus the LHS. Then

$$\alpha = (O) + ((P \star Q) \star O) - (P) - (Q).$$

We're trying to show that α is principal. Define

$$\beta := \operatorname{div}((P \star Q) * O)) - \operatorname{div}(P * Q).$$

Then β is principal (since principal divisors form an abelian group). Hence it is enough to show that $\alpha = \beta$. We compute:

$$\begin{aligned}\beta &= \operatorname{div}((P \star Q) * O)) - \operatorname{div}(P * Q)\\ &= ((P \star Q) + (O) + ((P \star Q) \star O)) - ((P) + (Q) + (P \star Q))\\ &= (O) + ((P \star Q) \star O) - (P) - (Q)\\ &= \alpha\end{aligned}$$

This shows that $P \oplus Q = ((P \star Q) \star O)$, as desired. □

Remark. The formula $P \oplus Q = (P \star Q) \star O$ may look a bit opaque, but in fact it arises in any situation where the goal is to define an abelian group $(A, \oplus)$, but the binary operation $\star$ on A given by the formula $x \star y = -x \oplus -y$ is easier to define than the actual group law. In particular, note that if $x, y \in A,$, then:

$$(x \star y) \star 0 = -(x \star y) \oplus 0 = -(x \star y) = -(-x \oplus -y) = x \oplus y.$$

4.3 Weierstrass elliptic curves

A special case of the planar elliptic curves considered in the previous section are the Weierstrass elliptic curves.

Definition 4.3.1. Given an element $\tilde{a} \in K^5$, the *Weierstrass cubic* associated to $\tilde{a}$ is defined as follows:

$$w_{\tilde{a}} = (y^2z + a_1xyz + a_3yz^2) - (x^3 + a_2x^2z + a_4xz^2 + a_6z^3),$$

and the *Weierstrass curve* associated to $\tilde{a}$, denoted $W(\tilde{a})$ is the projective vanishing set of this cubic. Explicitly:

$$W(\tilde{a}) := \{[x : y : z] \in \mathbb{P}^2 : y^2z + a_1xyz + a_3yz^2 = x^3 + a_2x^2z + a_4xz^2 + a_6z^3\}.$$

We'll sometimes refer to the elements of K^5 as *Weierstrass coefficient sequences* for emphasis.

Remark. One possible justification for the apparently strange numbering of the a_i's is this: the notation is chosen such that if we apply the homomorphism

$$x \mapsto x^2, y \mapsto y^3, z \mapsto 1$$

to the Weierstrass cubic $w_{\tilde{a}}$ and then re-homogenize it (causing all monomials to have a total degree of 6), then exponent of z ends up being equal to the subscript of a_i like so:

$$(y^6 + a_1x^2y^3z + a_3y^3z^3) - (x^6 + a_2x^4z^2 + a_4x^2z^4 + a_6z^6)$$

Proposition 4.3.2. Given a Weierstrass coefficient sequence $\tilde{a} \in \overline{K}^5$, the curve $W(\tilde{a})$ intersects the line at infinity at $O := [0 : 1 : 0]$ and only at this point.

Proof. Let $P = [x : y : z] \in \mathbb{P}^2$ be a point at infinity. Then $z = 0$. So $P = [x : y : 0]$. Thus:

$$P \in E \iff [x : y : 0] \in E \iff 0 = x^3 \iff x = 0.$$

Hence the points at infinity that lie on E are precisely those of the form $[0 : y : 0]$. But since $O = [0 : y : 0]$, hence this is the only point at infinity where an intersection occurs. $\square$

Since the only place where a Weierstrass curve intersects the line at infinity is at O, hence the study of Weierstrass curves can largely proceed by looking at the affine portion

$$W(\tilde{a}) \cap \overline{K}^2 = \{(x, y) \in \overline{K}^2 : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6\},$$

which is obtained by taking $z = 1$. The following definitions are consequently helpful.

Definition 4.3.3. Given a Weierstrass coefficient sequence $\tilde{a} \in K^5$, we have:

$$\begin{aligned} w_{\tilde{a}}(x, y) &:= w_{\tilde{a}}(x, y, 1) \\ w^x_{\tilde{a}}(x, y) &:= \frac{\partial}{\partial x} w_{\tilde{a}}(x, y) \\ w^y_{\tilde{a}}(x, y) &:= \frac{\partial}{\partial y} w_{\tilde{a}}(x, y) \end{aligned}$$

Proposition 4.3.4. Given a Weierstrass coefficient sequence $\tilde{a} \in K^5$, we have:

$$\begin{aligned} w_{\tilde{a}}(x, y) &= (y^2 + a_1xy + a_3y) - (x^3 + a_2x^2 + a_4x + a_6) \\ w^x_{\tilde{a}}(x, y) &= a_1y - (3x^2 + 2a_2x + a_4) \\ w^y_{\tilde{a}}(x, y) &= 2y + a_1x + a_3 \end{aligned}$$

Similarly, isogenies are usually defined between affine portions of Weierstrass curves only. By Proposition 3.7.2, this uniquely determines the corresponding morphisms of projective curves. However, actually homogenizing the formula for the affine portion of an isogeny to obtain the formula for a genuine isogeny between the corresponding projective curves seems to be non-trivial as far as I know.

Definition 4.3.5. A *elliptic Weierstrass curve* is a pair (E, O) where E is a smooth Weierstrass curve and $O = [0 : 1 : 0]$.

It's clear that if E is a smooth Weierstrass curve, then $(E, [0 : 1 : 0])$ is an elliptic Weierstrass curve. However, the question remains of how best to check smoothness. Showing that a Weierstrass curve is smooth by studying $w^x_{\tilde{a}}$ and $w^y_{\tilde{a}}$, is, in general, a non-trivial task. For this reason, it is helpful to have a notion of *discriminant* for a Weierstrass curve. It turns out that smoothness of the curve is equivalent to the discriminant being non-zero.

To define the discriminant, let us regard each a_i as function $K^5 \to K$. Further functions of interest with the same domain and codomain are specified below:

$$b_2 = a_1^2 + 4a_2, \qquad b_4 = 2a_4 + a_1a_3, \qquad b_6 = a_3^2 + 4a_6$$
$$b_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2$$
$$\Delta = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6$$

Definition 4.3.6. The *discriminant* of associated to a Weierstrass coefficient sequence $\tilde{a} \in K^5$ is the quantity $\Delta(\tilde{a})$ where Δ is defined above.

Proposition 4.3.7. Given a Weierstrass coefficient sequence $\tilde{a} \in K^5$, the curve $W(\tilde{a})$ is smooth if and only if the discriminant $\Delta(\tilde{a})$ is non-zero.

Proof. See [43, p.46]. □

It turns out that every elliptic curve is isomorphic to an elliptic Weierstrass curve.

Proposition 4.3.8. Let (E, O) denote an elliptic curve. Then there exists a Weierstrass coefficient sequence $\tilde{a} \in K^5$ together with an isomorphism

$$\tilde{\varphi} : (E, O) \to W(\tilde{a}).$$

Proof. This is Proposition 3.1 in Silverman [43]. □

We're now in a position to show that elliptic curves are abelian varieties with the group structure given in Definition 4.0.4. In particular:

Proposition 4.3.9. If (E, O) is an elliptic curve, then $(E, \oplus, O, -\square)$ is an abelian variety.

Proof. Let $\varphi : (E, O) \to W(\tilde{a})$ denote Weierstrass coordinates for (E, O). Since $\varphi(P \oplus Q) = \varphi(P) \oplus \varphi(Q)$ by Proposition 4.1.2, hence $P \oplus Q = \varphi^{-1}(\varphi(P) \oplus \varphi(Q))$. Thus the addition law on (E, O) is given as $\varphi^{-1} \circ \sigma \circ (\varphi \times \varphi)$ where $\sigma : W(\tilde{a}) \times W(\tilde{a}) \to W(\tilde{a})$ is the group law on $W(\tilde{a})$. But the group operations on a Weierstrass curve are given explicitly in Silverman under the heading of Group Law Algorithm 2.3 [43, p.53] and hence these are regular. So in particular, σ is regular. Hence the group law on (E, O), being a composite of regular maps, is regular. The proof that taking inverses is a regular map is similar. □

There's an issue with the discriminant, which is that it's unstable under isomorphisms of elliptic curves. To address this issue, we define the following further quantities associated with Weierstrass coefficient sequences.

$$c_4 = b_2^2 - 24b_4, \qquad c_6 = -b_2^3 + 36b_2b_4 - 216b_6, \qquad j = c_4^3/\Delta$$

The quantity j defined above is called the j-invariant of the curve. Unlike the discriminant, it's stable under isomorphism. This allows us to talk about the j-invariant $j(E)$ of an arbitrary elliptic curve E (whereas the discriminant only applies to a Weierstrass curve.) In fact, the j-invariant provides a complete characterization of elliptic curves up to $\overline{K}$-isomorphism.

Proposition 4.3.10. (a) Let E_0 and E_1 denote elliptic curves. Then E_0 and E_1 are isomorphic over $\overline{K}$ if and only if $j(E_0) = j(E_1)$.

(b) For all $t \in \overline{K}$, there exists a Weierstrass coefficient sequence $\tilde{a} \in \overline{K}^5$ such that $j(W(\tilde{a})) = t$.

Proof. This is a consequence of Proposition 1.4 in Silverman [43, p.45]. □

4.4 Other normal forms for Elliptic curves

The Weierstrass form is not the only way of describing elliptic curves. Especially relevant for the purposes of cryptography are *Montgomery curves*, which are degree 3 projective plane curves that are general enough to encompass many elliptic curves of interest (up to isomorphism), and *twisted Edwards curves*, which are degree 4 affine plane curves that are birationally equivalent to certain elliptic curves. It should be noted that twisted Edwards curves, being non-smooth plane curves, are not actually elliptic curves according to our definition (except when regarded up to birational equivalence).

Definition 4.4.1. Given a pair $(A, B) \in K^2$, the *affine Montgomery cubic* associated to (A, B) is defined as follows:

$$m_{A,B} = By^2 - (x^3 + Ax^2 + x),$$

and the *affine Montgomery curve* associated to (A, B), denoted $M(A, B)$ is the affine vanishing set of this cubic. Explicitly:

$$M(A, B) := \{(x, y) \in \mathbb{A}^2_K : By^2 = x^3 + Ax^2 + x\}.$$

The above definition can be moved to projective space in the obvious way to obtain an actual projective curve, called a projective Montgomery curve. When this curve is smooth, it's an elliptic curve as a consequence of the genus-degree formula. However, not every elliptic curve is isomorphic to a smooth Montgomery curve with a distinguished basepoint, as explained by Okeya et. al. [33], citing Montgomery [29]. For those that are, the addition and doubling formulas presented by Montgomery allow for faster arithmetic operations. Further efficiency benefits can be obtained by recasting into twisted Edwards form:

Definition 4.4.2. Given a pair $(a, b) \in K^2$, the *affine twisted Edwards quartic* associated to (a, b) is defined as follows:

$$e_{a,b} = ax^2 + y^2 - (1 + dx^2y^2),$$

and the *affine twisted Edwards curve* associated to (a, b), denoted $E(a, b)$ is the projective vanishing set of this quartic. Explicitly:

$$E(a, b) := \{(x, y) \in \mathbb{A}^2_K : ax^2 + y^2 = 1 + dx^2y^2\}.$$

Arithmetic on twisted Edwards curves is especially straightforward. With some caveats, the group law is given by:

$$(x_1, y_1) + (x_2, y_2) = \left(\frac{x_1y_2 + y_1x_2}{1 + dx_1x_2y_1y_2}, \frac{y_1y_2 - ax_1x_2}{1 - dx_1x_2y_1y_2}\right).$$

The identity element is $(0, 1)$ and negation is given by

$$-(x, y) = (x, -y).$$

Note the resemblance to the group law on a Pell conic discussed later in the thesis; I'm not sure if this leads to any further insights, but the comparison seems interesting. The interested reader is directed to the 2008 article in which twisted Edwards curves were originally proposed [6].

4.5 Separable and inseparable degree of isogenies

Let us now consider the separable and inseparable degrees of isogenies between elliptic curves. The following result will be helpful later.

Proposition 4.5.1. For all elliptic curves E, the set $E(\overline{K})$ has infinitely many points.

Proof. Since every elliptic curve is isomorphic to a Weierstrass curve, it is enough to show that every Weierstrass curve has infinitely many points. Hence it suffices to show that the affine part of any Weierstrass curve has infinitely many points. Let $\tilde{a} \in \overline{K}^5$ denote a sequence of Weierstrass coordinates. Define a function p as follows

$$\begin{aligned} W(\tilde{a}) \cap \overline{K}^2 &\xrightarrow{p} \overline{K} \\ (x, y) &\longmapsto x. \end{aligned}$$

It suffices to show that p is surjective. Consider $x \in \overline{K}$. Our goal is to show that there exists $y \in \overline{K}$ such that $(x, y) \in W(\tilde{a}) \cap \overline{K}^2$. That is, we're trying to show that

$$\exists y \in \overline{K} : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

Rearranging to make this into a polynomial equation in the variable y, it becomes

$$\exists y \in \overline{K} : y^2 + (a_1x + a_3)y + (-x^3 - a_2x^2 - a_4x - a_6),$$

which is true because $\overline{K}$ is algebraically closed. □

Some basic results about homomorphisms of groups will also be useful.

Proposition 4.5.2. Let $\varphi : G \to H$ denote a surjective morphism of abelian groups. Then for all $h \in H$, we have $\#\varphi^{-1}(h) = \#\ker(\varphi)$.

Remark. The above theorem is true even if one or both of the relevant kernels is infinite, as long as the function $\#\square$ is interpreted as returning a cardinal number.

Proof. Since φ is surjective, we can find $g \in G$ satisfying $\varphi(g) = h$. Define functions

$$\begin{tikzcd} G \arrow[r, "\square+g"] & G \\ x \arrow[r, mapsto] & x+g \end{tikzcd} \qquad \begin{tikzcd} G \arrow[r, "\square-g"] & G \\ x \arrow[r, mapsto] & x-g \end{tikzcd}$$

It is easy to see that these restrict to inverse functions

$$\begin{tikzcd} \ker(\varphi) \arrow[r, "\square+g"] & \varphi^{-1}(h) \\ x \arrow[r, mapsto] & x+g \end{tikzcd} \qquad \begin{tikzcd} \varphi^{-1}(h) \arrow[r, "\square-g"] & \ker(\varphi) \\ x \arrow[r, mapsto] & x-g \end{tikzcd}$$

Thus $\#\ker(\varphi) = \#\varphi^{-1}(h)$, as desired. □

Proposition 4.5.3. Let $\varphi : E_1 \to E_2$ denote a non-zero isogeny between elliptic curves. Then $\deg_s(\varphi) = \#\ker(\varphi)$.

Proof. Since φ is non-zero, hence it's non-constant. Hence by Proposition 3.8.7, for all but finitely many $Q \in E_2$ we have $\#\varphi^{-1}(Q) = \deg_s(\varphi)$. Since E_2 has infinitely many points by Proposition 4.5.1, hence there exists $Q \in E_2$ such that $\#\varphi^{-1}(Q) = \deg_s(\varphi)$. Since φ is a morphism of groups, hence by Proposition 4.5.2, we have that $\#\varphi^{-1}(Q) = \#\varphi^{-1}(O)$. Hence $\deg_s(\varphi) = \#\varphi^{-1}(O)$, as desired. □

Proposition 4.5.4. Consider two morphisms of abelian groups

$$G \xrightarrow{\varphi} H \xrightarrow{\psi} I.$$

If φ is surjective, then $\#\ker(\psi \circ \varphi) = \#\ker(\varphi) \cdot \#\ker(\psi)$.

Remark. The above theorem is true even if one or both of the relevant kernels is infinite.

Proof. Begin by computing

$$\begin{aligned}
\ker(\psi \circ \varphi) &= \varphi^{-1}(\psi^{-1}(0)) \\
&= \varphi^{-1}\left(\bigcup_{h \in \psi^{-1}(0)} \{h\} \right) \\
&= \bigcup_{h \in \psi^{-1}(0)} \big(\varphi^{-1}(h)\big)
\end{aligned}$$

It can be checked that this union is a disjoint union. Hence

$$\begin{aligned}
\#\ker(\psi \circ \varphi) &= \sum_{h \in \psi^{-1}(0)} \# \big(\varphi^{-1}(h)\big) \\
&= \sum_{h \in \psi^{-1}(0)} \# \big(\varphi^{-1}(h)\big) \\
&= \sum_{h \in \psi^{-1}(0)} \#\ker(\varphi) \qquad \text{by surjectivity of } \varphi \text{ and Proposition 4.5.2} \\
&= \#\ker(\varphi) \cdot \sum_{h \in \psi^{-1}(0)} 1 \\
&= \#\ker(\varphi) \cdot \#\ker(\psi)
\end{aligned}$$

This completes the proof. □

Proposition 4.5.5. Given two isogenies of elliptic curves

$$G \xrightarrow{\varphi} H \xrightarrow{\psi} I,$$

the separable and inseparable degree functions behave as follows:

$$\deg_s(\psi \circ \varphi) = \deg_s(\varphi) \cdot \deg_s(\psi), \qquad \deg_i(\psi \circ \varphi) = \deg_i(\varphi) \cdot \deg_i(\psi)$$

Proof. Let's first address the separable degree. There are two cases. If at least one of φ or ψ is constant, then so too is $\psi \circ \varphi$, and hence both sides of the formula are 0 and the equality holds. So assume neither φ nor ψ is constant. Then ψ is surjective by Proposition 3.8.1, and $\psi \circ \varphi$ is non-constant by Proposition 3.8.4. Hence we may argue as follows:

$$\begin{aligned}
\text{LHS} &= \deg_s(\psi \circ \varphi) \\
&= \#\ker(\psi \circ \varphi) && \text{by Proposition 4.5.3} \\
&= \#\ker(\varphi) \cdot \#\ker(\psi) && \text{by Proposition 4.5.4} \\
&= \deg_s(\varphi) \cdot \deg_s(\psi) && \text{by Proposition 3.8.7} \\
&= \text{RHS}
\end{aligned}$$

This establish the formula. Let's now address the inseparable degree. Once again if at least one of φ or ψ is constant, it's trivial. So assume neither φ nor ψ is constant, and infer that $\psi \circ \varphi$ is non-constant. Then by Proposition 1.4.4, we obtain:

$$\deg(\psi \circ \varphi) = \deg_s(\psi \circ \varphi)\deg_i(\psi \circ \varphi).$$

Hence using Proposition 3.8.8, we infer:

$$\deg(\varphi) \cdot \deg(\psi) = \deg_s(\varphi)\deg_s(\psi)\deg_i(\psi \circ \varphi).$$

Using Proposition 1.4.4 again, we obtain:

$$\deg_s(\varphi)\deg_i(\varphi) \cdot \deg_s(\psi)\deg_i(\psi) = \deg_s(\varphi)\deg_s(\psi)\deg_i(\psi \circ \varphi).$$

Cancelling, it follows that:

$$\deg_i(\varphi)\deg_i(\psi) = \deg_i(\psi \circ \varphi). \qquad \square$$

4.6 Ordinary versus supersingular elliptic curves

For each elliptic curve E, the isogenies $E \to E$ form an algebra called the *endomorphism algebra* of E, denoted $\text{End}(E)$. (This is usually called the endomorphism ring of E, but recall that, for us, rings are always commutative.) The zero element is the zero isogeny, and the multiplicative unit is the identity function. By the initiality of $\mathbb{Z}$ in the category of algebras, there is a unique algebra morphism $\mathbb{Z} \to \text{End}(E)$ denoted $[\square]$. It can be computed as follows:

$$[n](P) = [\underbrace{1 + \cdots + 1}_{n}](P) = \underbrace{[1](P) + \cdots + [1](P)}_{n} = \underbrace{P + \cdots + P}_{n}$$

The kernel of this map is denoted $E[n]$. Explicitly:

$$E[n] = \{P \in E : [n]P = 0\}.$$

The following proposition shows that in positive characteristic, there are two very different kinds of elliptic curves, called *ordinary* and *supersingular* elliptic curves respectively.

Proposition 4.6.1. Assume the ground field K has characteristic $p > 0$. For all elliptic curves (E, O), either

(a) (ordinary) for all positive integers r we have $E[p^r] = \mathbb{Z}/p^r$, or

(b) (supersingular) for all positive integers r we have $E[p^r] = 0$.

Proof. This is a consequence of Theorem 3.1 in Silverman [43, p.144], whom cites Duering [16]. □

Proposition 4.6.2. Assume the ground field K has characteristic $p > 0$. For all elliptic curves (E, O), we have

(a) E is ordinary if and only if $\mathrm{End}(E)$ is an order in an imaginary quadratic algebra

(b) E is supersingular if and only if $\mathrm{End}(E)$ an order in a quaternion algebra.

Proof. This is a consequence of Theorem 3.1 in Silverman [43, p.144] combined with the conclusion of Exercise 5.8 [43, p.154]. □

Hence the distinction between ordinary and supersingular amounts to whether the corresponding endomorphism ring is commutative:

Corollary 4.6.3. Assume the ground field K has characteristic $p > 0$. For all elliptic curves (E, O), we have

(a) E is ordinary if and only if $\mathrm{End}(E)$ is commutative.

(b) E is supersingular if and only if $\mathrm{End}(E)$ is non-commutative.

Non-commutativity of the endomorphism algebra of E when E is non-singular will play a role in our discussion of the SIDH protocol in the cryptography portion of the thesis.

4.7 Velu's Formula

Recall that elliptic curves don't have useful notion of sub-elliptic curve. One might be tempted to conclude that there is no meaningful notion of quotient elliptic curve, since for abelian groups, there's a natural bijective correspondence between subgroups and quotient groups. However, this conclusion would be overly hasty, and in particular, it is possible to take a quotient of an elliptic curve by a finite subgroup to obtain a new elliptic curve.

A very general comment is that quotients in algebraic geometry are typically nontrivial to construct, and their study leads to an involved body of knowledge called geometric invariant theory [31]. However, taking the quotient of a variety by the action of a finite group is simpler than the general problem. In particular, the quotient of a variety by a finite group of automorphisms is again a variety, according to Silverman [43, p.74], whom cites Mumford [32, §7]. A special case of this is the quotient of an elliptic curve by a finite subgroup, which turns out to be amenable to an especially elementary analysis.

We have shown that every elliptic curve E is isomorphic to an elliptic Weierstrass curve. Velu's formula [45] tells us how to take quotients of Weierstrass curves. In particular, it tells us how to construct, for any Weierstrass coefficient sequence $\tilde{a} \in K^5$ such that $W(\tilde{a})$ is smooth, and any finite subgroup $F \subseteq W(\tilde{a})$, a Weierstrass coefficient sequence $\tilde{A} \in K^5$ such that $W(\tilde{A})$ is smooth, together with a separable isogeny $\varphi : W(\tilde{a}) \to W(\tilde{A})$ whose kernel is precisely F. It makes sense to define $\tilde{a}/F$ to mean the coefficient sequence $\tilde{A}$, though this is not a standard notation.

A couple of remarks are in order. Firstly, it should be noted that Velu's formula only gives us the formula for the affine portion of the isogeny φ. By Proposition 3.7.2, this is enough to uniquely specify an isogeny between the full projective curves. Secondly, the actual presentation of Velu's article will differ slightly from our description here, the main difference being that Velu works with the function $g_{\tilde{a}}(x, y) := -w_{\tilde{a}}(x, y)$. We make the necessary changes here to accommodate this difference, ensuring that our t and u functions agree with Velu's and that the final coefficients are unchanged.

For the remainder of this subsection, let $\tilde{a} \in K^5$ denote a Weierstrass coefficient sequence, and let x and y denote the projection maps $W(\tilde{a}) \cap \overline{K}^2 \to \overline{K}$ acting on the affine portion of the corresponding curve. Now recall the following formulae from Proposition 4.3.4 for the derivatives of $w_{\tilde{a}}$, which can also be viewed as functions on $W(\tilde{a}) \cap \overline{K}^2 \to \overline{K}$, and define a further function $t_{\tilde{a}}$ with this same domain and codomain as follows:

$$\begin{aligned} w^x_{\tilde{a}} &= a_1 y - (3x^2 + 2a_2 x + a_4) \\ w^y_{\tilde{a}} &= 2y + a_1 x + a_3 \\ t_{\tilde{a}} &= \begin{cases} -w^x_{\tilde{a}} & Q \in W(\tilde{a})[2] \\ -2w^x_{\tilde{a}} + a_1 w^y_{\tilde{a}} = 6x^2 + b_2 x + b_4 & Q \notin W(\tilde{a})[2] \end{cases} \\ u_{\tilde{a}} &= (w^y_{\tilde{a}})^2 = 4x^3 + b_2 x^2 + 2b_4 x + b_6 \end{aligned}$$

Now proceed as follows:

- Let $F \subseteq W(\tilde{a})$ denote the finite subgroup that we wish to quotient out by, and note that $F \setminus \{O\}$ is the affine portion F; that is, $F \setminus \{O\} = F \cap \overline{K}^2$.
- Define $F_2 := F[2] \setminus \{O\}$ to consist of all elements of $F \setminus \{O\}$ of order 2.
- Fix a set $R \subseteq F \setminus F[2]$ satisfying $R \cup -R = F \setminus F[2]$ and $R \cap -R = \emptyset$.
- Define scalars as follows:

$$T_{\tilde{a},F} = \sum_{Q \in R \cup F_2} t_{\tilde{a}}(Q), \qquad W_{\tilde{a},F} = \sum_{Q \in R \cup F_2} (u_{\tilde{a}} + x_{\tilde{a}} t_{\tilde{a}})(Q)$$

Remarkably, the scalars $T_{\tilde{a},F}$ and $W_{\tilde{a},F}$ depend only only on the subgroup F and not on the decomposition R that was chosen. Essentially, Velu's formula is the statement that:

$$(\tilde{a}/F)_1 = a_1, \qquad (\tilde{a}/F)_3 = a_3$$

$$(\tilde{a}/F)_2 = a_2, \qquad (\tilde{a}/F)_4 = a_4 - 5T_{\alpha,F}, \qquad (\tilde{a}/F)_6 = a_6 - b_2 T_{\alpha,F} - 7W_{\alpha,F}.$$

His article also gives us (the affine portion of) an isogeny $(X, Y) : W(\tilde{a}) \to W(\tilde{a}/F)$ to the quotient curve. The isogeny (X, Y) turns out to be separable and its kernel is F. Explicitly, it's given by

$$X = x + \sum_{Q \in R \cup F_2} \left(\frac{t_{\tilde{a}}(Q)}{x - x(Q)} + \frac{u_{\tilde{a}}(Q)}{(x - x(Q))^2} \right)$$

$$Y = y - \sum_{Q \in R \cup F_2} \left(u_{\tilde{a}}(Q) \frac{2y + a_1 x + a_3}{(x - x(Q))^3} + t_{\tilde{a}}(Q) \frac{a_1(x - x(Q)) + y - y(Q)}{(x - x(Q))^2} + \frac{a_1 u_{\tilde{a}}(Q) - w^x(Q) w^y(Q)}{(x - x(Q))^2} \right)$$

Consider the elliptic curve

$$W(\tilde{a}) : y^2 + xy + y = x^3 - x^2 - 3x + 3.$$

This has Weierstrass coordinates given by he projection functions

$$\begin{aligned} W(\tilde{a}) \cap \overline{K}^2 &\xrightarrow{x} \overline{K} \\ (P_0, P_1) &\longmapsto P_0 \end{aligned} \qquad\qquad \begin{aligned} W(\tilde{a}) \cap \overline{K}^2 &\xrightarrow{y} \overline{K} \\ (P_0, P_1) &\longmapsto P_1 \end{aligned}$$

Let F denote the finite subgroup of E with 7 elements generated by $Q = (1, 0)$. Explicitly, its elements are:

$$0Q = O \qquad Q = (1, 0) \qquad 2Q = (-1, -2)$$

$$3Q = (3, -6) \qquad 4Q = (3, 2) \qquad 5Q = (-1, 2) \qquad 6Q = (1, -2).$$

Then $F[2] = \{O\}$, so $F_2 = \emptyset$. We choose a decomposition $R = \{Q, 2Q, 3Q\}$. The resulting numbers are displayed below.

$$w^x(Q) = 2, \quad w^x(2Q) = -4, \quad w^x(3Q) = -24$$

$$w^y(Q) = 2, \quad w^y(2Q) = -4, \quad w^y(3Q) = -8$$

$$t(Q) = -2, \quad t(2Q) = 4, \quad t(3Q) = 40$$

$$u(Q) = 4, \quad t(2Q) = 16, \quad t(3Q) = 64$$

$$T = 42, \qquad W = 198$$

Note that our value of $w^x(Q)$ is different from (the negative of) Velu's answer; this is a misprint in Velu's article. We find that:

$$(\tilde{a}/F)_1 = 1, \qquad (\tilde{a}/F)_3 = 1$$

$$(\tilde{a}/F)_2 = -1, \qquad (\tilde{a}/F)_4 = -213, \qquad (\tilde{a}/F)_6 = -1257.$$

So the (affine portion) of the quotient curve $W(\tilde{a}/F)$ is

$$y^2 + xy + y = x^3 - x^2 - 213x - 1257.$$

The code used to perform the above computations can be found in Appendix A.

4.8 The group of $\mathbb{Q}$-rational points of an elliptic curve

The following group-theoretic preliminaries will be helpful.

Definition 4.8.1. Whenever X is an abelian group, write $X[n]$ for the subgroup of n-torsion elements

$$X[n] = \{x \in X : nx = 0\},$$

and write T_X for the subgroup of torsion elements:

$$T_X = \bigcup_{n \in \mathbb{N}} X[n].$$

Proposition 4.8.2 (Structure Theorem For Finitely-Generated Abelian Groups)**.** If X is a finitely-generated abelian group, then there is a unique natural number r such that $X \cong T_X \oplus \mathbb{Z}^r$.

Proof. Here's a sketch of the proof. It is easy to show that for all abelian groups X, the group X/T_X is torsion-free. Much harder to prove is that for each finitely-generated abelian group X, there exists an isomorphism $X \to T_X \oplus X/T_X$, but this is also true. Now recall that over a principal ideal domain, every finitely-generated torsion-free module is free. Hence noting that $\mathbb{Z}$ is a PID, we deduce that if X is finitely-generated, then X/T_X is free. □

Definition 4.8.3. The number r in the above result is called the *rank* of X.

Returning to the topic of elliptic curves, Mordell's Theorem (below) allows us to apply the above analysis to the group $E(\mathbb{Q})$ for any elliptic curve E defined over $\mathbb{Q}$.

Proposition 4.8.4 (Mordell's Theorem; 1922)**.** Let E denote an elliptic curve defined over the rational numbers. Then the group $E(\mathbb{Q})$ is finitely-generated, and the group $T_{E(\mathbb{Q})}$ is finite.

Proof. This is proved in [30]. □

More than fifty years later, the possible torsion subgroups of $E(\mathbb{Q})$ were classified:

Proposition 4.8.5 (Mazur's Theorem; 1977)**.** Let E denote an elliptic curve defined over the rational numbers. Then either there exists $n \in \{1, \dots, 12\} \setminus \{11\}$ such that $T_{E(\mathbb{Q})} \cong \mathbb{Z}/n$, or there exists $m \in \{1, \dots, 4\}$ such that $T_{E(\mathbb{Q})} \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2m\mathbb{Z}$.

Proof. This is proved in [25]. □

In contrast, the rank of $E(\mathbb{Q})$ is not well-understood; however, some interesting asymptotic results have recently been discovered. Particularly noteworthy is:

Proposition 4.8.6 (Bhargava, Shankar; 2010-2012)**.** The average rank of all elliptic curves over $\mathbb{Q}$ is less than 1.

Proof. This is proved in [7]. □

Manjul Bhargava received the 2014 Fields Medal for the work leading up to this result.

4.9 The cardinality of $E(\mathbb{F}_q)$

For the remainder of this section, assume that p is a prime number and that $q = p^n$ is a prime power. If E is an elliptic curve defined over $\mathbb{F}_q$, the group $E(\mathbb{F}_q)$, being a subset of $\mathbb{P}^2(\mathbb{F}_q)$, is clearly finite, and the question arises of how to find it's cardinality. Hasse's theorem doesn't give us a direct answer, but provides instead a relatively tight bound. Given $a, b \in \mathbb{R}$ and $r \in \mathbb{R}_{\geq 0}$, let's $a \overset{r}{=\!=} b$ to mean $|a - b| \leq r$, i.e. the distance between a and b is at most r. In this notation, we have:

Proposition 4.9.1 (Hasse; 1933)**.** For any elliptic curve E defined over $\mathbb{F}_q$, we have

$$\#E(\mathbb{F}_q) \overset{2\sqrt{q}}{=\!=\!=} q + 1.$$

Proof. This is Theorem 1.1 in Silverman [43, p.138]. □

To get an exact value for the cardinality of $\#E(\mathbb{F}_q)$, one can use Schoof's algorithm. The ideas behind the algorithm are non-trivial, but the interested reader is directed Schoof's original article [38] and to a more recent survey by the same author [39]. A good preparatory exercise is to solve Exercise 3.7 in Silverman [43, p.105] which covers many of the needed preliminaries. The algorithm presented in the original article has $O(\log^9 q)$ time-complexity [38], making it infeasible for large values of q, however the complexity but can be improved by e.g. using efficient integer and polynomial multiplication algorithms. Building on Schoof's work, the SEA (Schoof-Elkies-Atkin) algorithm provides a feasible solution to the problem finding the number of points on elliptic curves defined over $\mathbb{F}_q$ for very large values of q. A good introduction is Atkin's original article [3].

Chapter 5

Pointed Conics

The material in this chapter is largely an elaboration on the ideas put forward in Shirali's article *Groups associated with conics* [41], and the comparison with elliptic curves described in Lemmermeyer's survey article on the similarities between Pell conics and elliptic curves [23].

In this chapter, assume K is a field, not necessarily perfect, of characteristic distinct from 2. In what follows, we'll identify the sequence $(x_1, \dots, x_n)$ with the column vector

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Let's also agree that if a and b are elements of the same K-module, then we'll write $a \sim b$ to mean that there exists a non-zero scalar $k \in K$ such that $ak = b$.

5.1 Conics and the large matrix

Definition 5.1.1. Given a sequence $\tilde{a} = (A, B, C, D, E, F) \in K^6$, the *bivariate quadratic polynomial* associated to $\tilde{a}$ is defined as follows:

$$Q_{\tilde{a}} = Ax^2 + Bxy + Cy^2 + Dx + Ey + F.$$

The *bivariate quadratic form* associated to $\tilde{a}$ is

$$q_{\tilde{a}} = Ax^2 + Bxy + Cy^2.$$

The *large matrix* associated to $\tilde{a}$ is

$$M_{\tilde{a}} = \begin{bmatrix} 2A & B & D \\ B & 2C & E \\ D & E & 2F \end{bmatrix}.$$

The *small matrix* associated to $\tilde{a}$ is

$$m_{\tilde{a}} = \begin{bmatrix} 2A & B \\ B & 2C \end{bmatrix}.$$

The relationship between the polynomials and the matrices defined above is summarized in the next result.

Proposition 5.1.2. Given a sequence $\tilde{a} = (A, B, C, D, E, F) \in K^6$, we have:

$$Q_{\tilde{a}}(x, y) = \frac{1}{2}(x, y, 1)^T M_{\tilde{a}}(x, y, 1), \qquad q_{\tilde{a}}(x, y) = \frac{1}{2}(x, y)^T m_{\tilde{a}}(x, y).$$

Proof. This is readily verified by hand or using a computer algebra package. □

It is helpful to have explicit formulae for the determinants of the above matrices:

Proposition 5.1.3. Given a sequence $\tilde{a} = (A, B, C, D, E, F) \in K^6$, we have:

$$\det(M_{\tilde{a}}) = 2(4ACF - AE^2 - B^2F + BDE - CD^2), \qquad \det(m_{\tilde{a}}) = 4AC - B^2.$$

Proof. The computation is most easily performed by cofactor expansion. The details are left to the reader. □

Proposition 5.1.4. Given a coefficient sequence $\tilde{a} \in K^6$, if the corresponding large matrix $M_{\tilde{a}}$ has non-zero determinant, then the ideal $(\overline{Q}_{\tilde{a}}) \subseteq \overline{K}[\pi_1, \pi_2]$ is prime.

Proof. Assume toward a contradiction that $(\overline{Q}_{\tilde{a}})$ is non-prime. Then $\overline{Q}_{\tilde{a}}$ is reducible. So there exist scalars $a, b, c, d, e, f \in \overline{K}$ satisfying

$$(ax + by + c)(dx + ey + f) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F.$$

Expanding the left-hand-side and equating coefficients, we deduce that:

$$ad = A, \qquad ae + bd = B, \qquad be = C,$$

$$af + cd = D, \qquad bf + ce = E, \qquad cf = F.$$

We compute:

$$\begin{aligned}\det(M_{\tilde{a}}) &\sim 4ACF - AE^2 - B^2F + BDE - CD^2 \\ &= 4adbecf - ad(bf+ce)^2 - (ae+bd)^2cf + (ae+bd)(af+cd)(bf+ce) - be(af+cd)^2 \\ &= 0\end{aligned}$$

Hence $\det(M_{\tilde{a}}) = 0$. But this contradicts that $\det(M_{\tilde{a}})$ is non-zero. □

Definition 5.1.5. A set $X \subseteq \mathbb{A}^2_K$ is said to be a *conic* if and only if there exists a coefficient sequence $\tilde{a} \in K^6$ such that $X = V(Q_{\tilde{a}})$ and $M_{\tilde{a}}$ is a nonsingular matrix.

Corollary 5.1.6. Every conic is an affine variety.

Proposition 5.1.7. Given coefficient sequences $\tilde{a}, \tilde{b} \in K^6$ such that $\det(M_{\tilde{a}}), \det(M_{\tilde{b}}) \neq 0$, from $V(Q_{\tilde{a}}) = V(Q_{\tilde{b}})$ we can infer the following:

$$Q_{\tilde{a}} \sim Q_{\tilde{b}}, \qquad q_{\tilde{a}} \sim q_{\tilde{b}}$$

$$M_{\tilde{a}} \sim M_{\tilde{b}}, \qquad m_{\tilde{a}} \sim m_{\tilde{b}}$$

Proof. From $V(Q_{\tilde{a}}) = V(Q_{\tilde{b}})$ we infer $I(V(Q_{\tilde{a}})) = I(V(Q_{\tilde{b}}))$. By Proposition 3.1.7, we deduce $\sqrt{(Q_{\tilde{a}})} = \sqrt{(Q_{\tilde{b}})}$. But by Proposition 5.1.4 this means that $(Q_{\tilde{a}}) = (Q_{\tilde{b}})$. Since the only units in $K[x, y]$ are the elements of K^*, it follows that $Q_{\tilde{a}} \sim Q_{\tilde{b}}$. Equating coefficients, we find that $\tilde{a} \sim \tilde{b}$. From this all four identities easily follow. □

Unlike with elliptic curves, where geometric irreducibility and smoothness had to be checked separately, for conics the situation is simpler.

Proposition 5.1.8. Every conic is smooth.

Proof. Consider fixed but arbitrary $\tilde{a} = (A, B, C, D, E, F) \in K^6$ and assume that $\det(M_{\tilde{a}})$ is non-zero. Assume toward a contradiction that there exists $(x, y) \in \mathbb{A}^2_K$ satisfying

$$Q_{\tilde{a}}(x, y) = 0, \qquad Q^x_{\tilde{a}}(x, y) = 0, \qquad Q^y_{\tilde{a}}(x, y).$$

This means that:

$$\begin{aligned}2Ax + By + D &= 0 \\ Bx + 2Cy + E &= 0 \\ q_{\tilde{a}}(x, y) + Dx + Ey + F &= 0\end{aligned}$$

Substituting the values of D and E we obtain from the two linear equations above into the quadratic equation, we obtain:

$$Ax^2 + Bxy + Cy^2 + (-2Ax - By)x + (-Bx - 2Cy)y + F = 0,$$

which can be rearranged to

$$Ax^2 + Bxy + Cy^2 = F.$$

In other words, $q_{\tilde{a}}(x, y) = F$. Hence our system of equations becomes

$$\begin{aligned} 2Ax + By + D &= 0 \\ Bx + 2Cy + E &= 0 \\ Dx + Ey + 2F &= 0 \end{aligned}$$

Or in other words, $M_{\tilde{a}}(x, y, 1) = 0$. But since the determinant of $M_{\tilde{a}}$ is non-zero, we deduce that $(x, y, 1) = M_{\tilde{a}}^{-1}(0, 0, 0)$ Thus $1 = 0$, a contradiction. □

5.2 The group law on a pointed conic

Definition 5.2.1. A *pointed conic* is a pair (X, N) such that X is a conic and $N \in X$ is a K-rational point.

Just like elliptic curves come naturally equipped with a group law, so too does each pointed conic become a group in a natural way.

Definition 5.2.2. Let (X, N) denote a pointed conic. Given $P, Q \in C$, we define a line $P * Q$ as follows:

(a) If $P \neq Q$, then $P * Q$ is the unique line that contains both P and Q

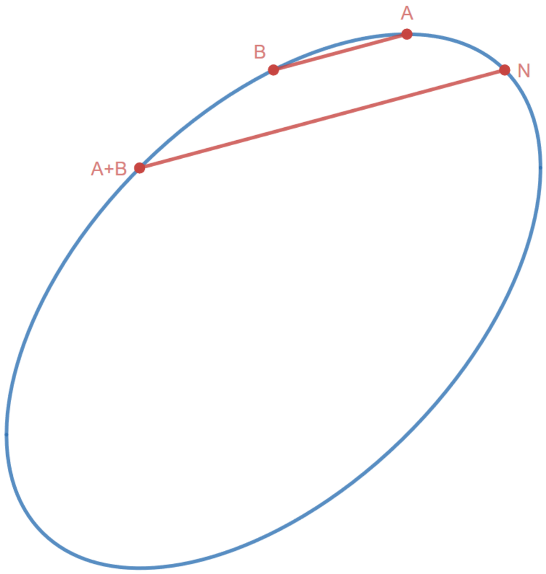
(b) If $P = Q$, then $P * Q$ is the tangent line to C at P.

Definition 5.2.3. Let (X, N) denote a pointed conic. Given $P, Q \in X$, we define $P \oplus Q$ by the formula:

$$\mathrm{div}_C(P * Q) = (N) + (P \oplus Q).$$

Geometrically, this involves drawing a line through P and Q, then translating it so as to go through the distinguished point N. The reader is directed to Shirali's recent

Figure 5.1: The group law on a pointed conic



article [41] for a proof that the group axioms are satisfied. Shirali goes further, using the classification of (non-degenerate) conics over the real line to describe the possible groups obtained this way (over $\mathbb{R}$) explicitly. In particular, since the group law is defined by drawing lines and translating them around, it's preserved under affine isomorphism, and thus problem can be simplified.

Definition 5.2.4. Let $X \subseteq \mathbb{A}^2_K$ denote a conic and let $\tilde{a} \in K^6$ satisfy $X = V(Q_{\tilde{a}})$. Then X is *parabolic* if and only if $\det(m_{\tilde{a}}) = 0$ and *non-parabolic* if and only if $\det(m_{\tilde{a}}) \neq 0$. If the ground field K is an ordered field, then we define that X is *elliptic* if and only if $\det(m_{\tilde{a}}) > 0$ and that X is a *hyperbolic* if and only if $\det(m_{\tilde{a}}) < 0$.

Remark. By Proposition 5.1.7, the small matrix of a conic is determined up to multiplication by an element of K^*. This tells us that the determinant of the small matrix is determined up to multiplication by a non-zero perfect square, and hence that the above definitions make sense.

It is a simple algebra check to see that affine isomorphisms preserve the above classification. Shirali's article exploits the classical result that, over the real line, every (non-degenerate) conic is affinely isomorphic to one of

$$x^2 + y^2 = 1, \qquad y = x^2, \qquad xy = 1,$$

depending on whether it's elliptic, parabolic, or hyperbolic, in order to find the groups explicitly in each of these cases. He concludes that, over the real line, we have:

Ellipse $\approx S^1 \approx \mathbb{R}/\mathbb{Z}$

Parabola $\approx (\mathbb{R}, +)$

Hyperbola $\approx (\mathbb{R}^*, \times) \cong (\mathbb{R}, +) \times \mathbb{Z}/2$

5.3 Arrows between pointed conics

In light of Shirali's results, a natural question to ask is: 'what is the correct notion of morphism between pointed conics?' Three possibilities suggest themselves, listed below in order of increasing generality:

affine homomorphism	basepoint preserving affine transformation
regular homomorphism	group-structure preserving regular mapping
isogeny	basepoint preserving regular map

Over the real line, all three notions listed above are distinct notions. To see that not every regular homomorphism is affine, consider the function:

$$x^2 + y^2 = 1, (1,0) \xrightarrow{[2]} x^2 + y^2 = 1, (1,0)$$
$$(x, y) \longmapsto (x^2 - y^2, 2xy).$$

This is not an affine homomorphism because $[2](1,0) = [2](-1,0)$. To see that not every isogeny is a regular homomorphism, consider the function

$$y^2 = x^2 + 1, (0,1) \xrightarrow{\gamma} y = x^2 + 1, (0,1)$$
$$(x, y) \longmapsto (x, y^2).$$

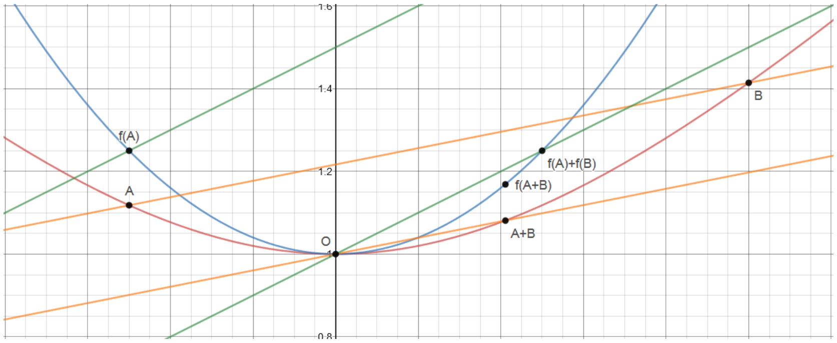
Figure 5.2 shows by way of example that γ is not a homomorphism of groups.

Nonetheless, it turns out that for certain kinds of conics, every isogeny is automatically a regular homomorphism, mirroring the situation with elliptic curves.

Proposition 5.3.1. Let K be a field of characteristic not 2. Then every isogeny between non-parabolic pointed conics is a regular homomorphism.

As far as I know, this is a new result. To prove it, let us first note that since isogenies are regular mappings, hence it only needs to be shown that every isogeny is a group

Figure 5.2: The isogeny $\gamma : (x, y) \mapsto (x, y^2)$ is not a group homomorphism



homomorphism. Our strategy will be to prove this for self-isogenies of the hyperbola $\{(x, y) \in \overline{K}^2 : xy = 1\}$, and then reduce the general case to this special case.

Proposition 5.3.2. Let K be a field of characteristic not 2. Then every self-isogeny of the pointed conic $(V(\pi_1\pi_2 - 1), (1, 1))$ is a group homomorphism.

Proof. We'll give a somewhat informal proof; a more rigorous approach would require the development of a theory of quasi-affine varieties, which is beyond the scope of this thesis. Elementary algebra shows that the functions

$$\begin{aligned} K^* &\xrightarrow{\alpha} V(\pi_1\pi_2 - 1) \\ x &\longmapsto (x, x^{-1}) \end{aligned} \qquad\qquad \begin{aligned} V(\pi_1\pi_2 - 1) &\xrightarrow{\beta} K^* \\ (x, y) &\longmapsto x \end{aligned}$$

are inverses. I claim that they are group homomorphisms. It suffices to show that α is a group homomorphism. Consider $x, y \in K^*$. Our goal is to show that $\alpha(xy) = \alpha(x) \oplus \alpha(y)$. There are two cases. The first case is where $\alpha(x) \neq \alpha(y)$. It suffices to show that the line through $\alpha(x)$ and $\alpha(y)$ is parallel to the line through $(1, 1)$ and $\alpha(xy)$. That is, it is enough to shos that

$$\alpha(y) - \alpha(x) \sim \alpha(xy) - (1, 1).$$

In other words, we're trying to show that

$$(y - x, y^{-1} - x^{-1}) \sim (xy - 1, x^{-1}y^{-1} - 1).$$

But this is equivalent to

$$(x^{-1} - y^{-1}, y - x) \bullet (xy - 1, x^{-1}y^{-1} - 1) = 0.$$

But this is easily verified by elementary algebra. The second case is where $\alpha(x) = \alpha(y)$. The goal is to show that the normal to the curve at $\alpha(x)$ is perpendicular to the line through $(1, 1)$ and $\alpha(x^2)$. That is, we're trying to show

$$(y, x) \bullet (x^2 - 1, x^{-2} - 1) = 0.$$

This is equivalent to showing that

$$((x^2 - 1)(xy - 1))/x = 0$$

as can be seen by expanding out both left-hand-sides. But since $(x, y) \in V(\pi_1\pi_2 - 1)$, the above statement is clearly true. Hence α is a homomorphism of groups, and thus so too is β.

Now consider an isogeny $\varphi : V(\pi_1\pi_2 - 1) \to V(\pi_1\pi_2 - 1)$. The mapping $\psi := \beta \circ \varphi\alpha : K^* \to K^*$ is a self-map of the quasi-affine variety $K^* = \mathbb{A}^1 \setminus \{0\}$. Hence ψ is of the form $\psi(x) = Ax^n$ for appropriate choices of $A \in K$ and $n \in \mathbb{Z}$. Now compute:

$$\begin{aligned} \psi(1) &= \beta(\varphi(\alpha(1))) \\ &= \beta(\varphi(1, 1)) \\ &= \beta(1, 1) && \text{because } \varphi \text{ is an isogeny} \\ &= 1 \end{aligned}$$

Thus $\psi(1) = 1$. In other words, $A \cdot 1^n = 1$. Hence $A = 1$. Hence ψ is given by the formula $\psi(x) = x^n$. It follows that ψ is a group homomorphism. Thus $\alpha \circ \psi \circ \beta$ is a group homomorphism. But since α and β are inverse functions, this composite equals φ. We deduce that φ is a group homomorphism, as desired. □

Next, we'll reduce the general case to the special case.

Proposition 5.3.3. Let K be a field of characteristic not 2 in which every element has a square root. Then any non-parabolic conic over K is affinely isomorphic to $xy = 1$.

Proof. Consider the conic $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$. Without loss of generality, assume $A \neq 0$. Complete the square on the homogeneous part by treating A, By and Cy^2 as the coefficients. The coefficient of x is unchanged; by hypothesis, the new coefficient of y (call it C') is non-zero (otherwise the small matrix would have determinant 0). Now complete the square two more times, once with x, once with y, to get the quadratic into the form $Ax^2 - C'y^2 + F' = 0$ for some $F' \in K$. We know F' is non-zero (otherwise the large matrix would have determinant 0). So divide through by F' to get into the form $\alpha x^2 - \beta y^2 = 1$. Now use the affine isomorphism $\alpha x^2 - \beta y^2 \to xy = 1$ given by $(x, y) \mapsto (\sqrt{\alpha}x + \sqrt{\beta}y, \sqrt{\alpha}x - \sqrt{\beta}y)$ to complete the proof. □

Putting these results together demonstrates Proposition 5.3.1 for any field of characteristic distinct from 2. Note that the base field does not have to be algebraically closed, since we can move to the algebraic closure to show that the isogeny is a homomorphism, and thereby conclude that it's a homomorphism over the ground field.

It should be noted that considering vanishing sets in an algebraic-closure is essential for Proposition 5.3.1. For instance, note that the function

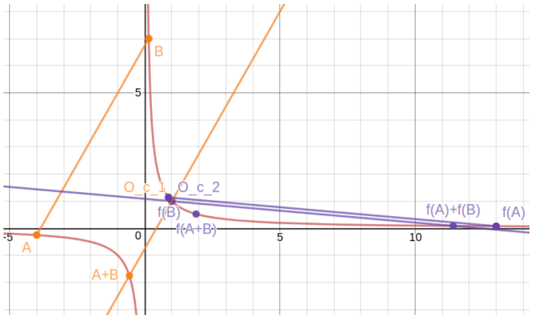
$$\begin{aligned} \mathbb{R} \setminus \{0\} &\xrightarrow{\varphi} \mathbb{R} \setminus \{0\} \\ x &\longmapsto x^2 - x + 1 \end{aligned}$$

satisfies $\varphi(1) = 1$, but is not a group homomorphism. For instance, we have $\varphi(2 \cdot 2) = 13$, but $\varphi(2)\varphi(2) = 9$. This illustrates the importance of working over an algebraically-closed field, for the proof above would fail without this assumption. Indeed, the function

$$\begin{aligned} V^2(\pi_1\pi_2 - 1) &\xrightarrow{\delta} V^2(\pi_1\pi_2 - 1) \\ (x, y) &\longmapsto (x^2 - x + 1, (x^2 - x + 1)^{-1}) \end{aligned}$$

is easily seen to preserve $(1, 1)$, yet is not a group homomorphism; see Figure 5.3.

Figure 5.3: Failure of the 'isogeny' $\delta : (x, y) \mapsto (x^2 - x + 1, (x^2 - x + 1)^{-1})$ to preserve the group law on the hyperbola.



5.4 Pell conics

Definition 5.4.1. Given a scalar $d \in K^*$ the *Pell conic* associated to d is the affine curve

$$C(d) = V^2(\pi_1^2 - d\pi_2^2 - 1) = \{(x, y) \in \overline{K}^2 : x^2 - dy^2 = 1\}.$$

Since the determinant of the corresponding large matrix is d, which is non-zero, we deduce that every Pell conic is indeed a conic. Since the determinant of the corresponding small matrix is $-d$, which is also non-zero, we deduce that every Pell conic is non-parabolic. It's also easy to see that the point $(1, 0)$ lies on $C(d)$ for all d. Taking this as our identity element, it follows that every Pell conic becomes a group. The law can be described explicitly as follows:

Proposition 5.4.2. For all $d \in K^*$, the group law on the Pell conic $C(d)$ is given by

$$(x_1, y_1) \oplus (x_2, y_2) = (x_1x_2 + dy_1y_2, x_1y_2 + y_1x_2).$$

Proof. Let us begin by checking well-definedness of the law. Assume:

$$x_1^2 - dy_1^2 = 1, \qquad x_2^2 - dy_2^2 = 1.$$

The goal is to show that

$$(x_1x_2 + dy_1y_2)^2 - d(x_1y_2 + y_1x_2)^2 = 1,$$

which is equivalent to proving

$$(dy_1^2 - x_1^2)(dy_2^2 - x_2^2) = 1,$$

as can be seen by expanding out both left-hand sides and noting that they're equal. Hence it's enough to prove that $(-1)(-1) = 1$, which is clearly true.

It remains to show that the line $(x_1, y_1) * (x_2, y_2)$ given in Definition 5.2.2 contains the point $(x_1x_2 + dy_1y_2, x_1y_2 + y_1x_2)$. There are two cases. For the first case, we assume $(x_1, y_1) \neq (x_2, y_2)$. To show that the relevant lines are parallel, it's enough to show that

$$(x_2 - x_1, y_2 - y_1) \sim (x_1x_2 + dy_1y_2 - 1, x_1y_2 + y_1x_2).$$

But this is equivalent to the statement

$$(y_1 - y_2, x_2 - x_1) \bullet (x_1x_2 + dy_1y_2 - 1, x_1y_2 + y_1x_2) = 0.$$

Expanding this out and using the formulae $x_1^2 - dy_1^2 = 1$ and $x_2^2 - dy_2^2 = 1$ proves the claim. For the other case, assume $(x_1, y_1) = (x_2, y_2)$. Define $(x, y) := (x_1, y_1) = (x_2, y_2)$. We must show that the tangent to C_d at (x, y) is parallel to the line $(x, y) * (x, y)$. It's enough to show that

$$(2dy, 2x) \sim (x^2 + dy^2 - 1, 2xy).$$

But this is equivalent to the statement that

$$(2x, -2dy) \bullet (x^2 + dy^2 - 1, 2xy) = 0.$$

So it's enough to show that $2x(x^2 - dy^2 - 1) = 0$. But using that $(x, y) \in C(d)$, this follows easily. □

Unsurprisingly, under certain conditions on d, the Pell conic $C(d)$ is isomorphic to the group of units of K.

Proposition 5.4.3. Suppose $d \in K$ has a square root in K. Then the function

$$\begin{aligned} C(d)(K) &\xrightarrow{\varphi_d} K^* \\ (x, y) &\longmapsto x - y\sqrt{d} \end{aligned}$$

is a morphism of groups.

Proof. Let's show that φ_d is well-defined. Suppose $x^2 - dy^2 = 1$. Our goal is to prove that $x - y\sqrt{d}$ is a unit. It suffices to show that $(x - y\sqrt{d})(x + y\sqrt{d}) = 1$. But this is clear.

Let us now show that φ_d is a morphism of groups. Assume

$$x_1^2 - dy_1^2 = 1, \qquad x_2^2 - dy_2^2 = 1.$$

Our goal is to show that

$$\varphi(x_1x_2 + dy_1y_2, x_1y_2 + y_1x_2) = \varphi(x_1, y_1)\varphi(x_2, y_2).$$

That is, we're trying to show that

$$(x_1x_2 + dy_1y_2) - (x_1y_2 + y_1x_2)\sqrt{d} = (x_1 - y_1\sqrt{d})(x_2 - y_2\sqrt{d}).$$

But this can be seen immediately by expanding out the right-hand side. □

Proposition 5.4.4. If K is not of characteristic 2, and if $d \in K$ has the property that $1/\sqrt{d}$ exists, then the function

$$\begin{aligned} K^* &\xrightarrow{\psi_d} C(d)(K) \\ t &\longmapsto \left(\frac{t^{-1}+t}{2}, \frac{t^{-1}-t}{2\sqrt{d}}\right) \end{aligned}$$

is an inverse to φ_d, and these are both consequently isomorphisms of groups.

Proof. Consider $t \in K^*$. We have:

$$\begin{aligned} &\varphi_d(\psi_d(t)) \\ &= \varphi_d\left(\frac{t^{-1}+t}{2}, \frac{t^{-1}-t}{2\sqrt{d}}\right) \\ &= \frac{t^{-1}+t}{2} - \frac{t^{-1}-t}{2\sqrt{d}}\sqrt{d} \\ &= \frac{t^{-1}+t}{2} - \frac{t^{-1}-t}{2} \\ &= t \end{aligned}$$

Consider $(x, y) \in C(d)(K)$. Observe that

$$(x - y\sqrt{d})^{-1} = \frac{1}{x - y\sqrt{d}} = \frac{x + y\sqrt{d}}{x^2 - y^2d} = x + y\sqrt{d}$$

Thus, we have:

$$
\begin{aligned}
&\psi_d(\varphi_d(x,y)) \\
&= \psi_d(x - y\sqrt{d}) \\
&= \left(\frac{(x - y\sqrt{d})^{-1} + (x - y\sqrt{d})}{2}, \frac{(x - y\sqrt{d})^{-1} - (x - y\sqrt{d})}{2\sqrt{d}} \right) \\
&= \left(\frac{(x + y\sqrt{d}) + (x - y\sqrt{d})}{2}, \frac{(x + y\sqrt{d}) - (x - y\sqrt{d})}{2\sqrt{d}} \right) \\
&= (x, y)
\end{aligned}
$$

This completes the proof. □

5.5 The rational unit circle

By Proposition 5.4.4, the rational unit hyperbola

$$C(1)(\mathbb{Q}) = \{(x,y) \in \mathbb{Q}^2 : x^2 - y^2 = 1\}$$

is isomorphic to $\mathbb{Q}^*$.

On the other hand, since -1 has no square root in $\mathbb{Q}$, the rational unit circle

$$C(-1)(\mathbb{Q}) = \{(x,y) \in \mathbb{Q}^2 : x^2 + y^2 = 1\}$$

is not subject to Proposition 5.4.4. Note that the set $C(-1)(\mathbb{Q})$ is closely related to the Pythogorean triples. In particular, given a Pythagorean triple (a, b, c), we can divide through the equation $a^2 + b^2 = c^2$ by c^2 to show that $(a/c, b/c)$ is an element of $C(-1)(\mathbb{Q})$ Conversely, if $(x, y) \in C(-1)(\mathbb{Q})$, then if $\lambda \in \mathbb{Z}$ is a common multiple of the denominators of x and y, we find that $(\lambda x, \lambda y, \lambda)$ is a Pythagorean triple. The group structure of the rational unit circle is known explicitly.

Proposition 5.5.1 (Tan, 1996)**.** Letting $\mathbb{P}$ denote the set of prime numbers, we have:

$$C(-1)(\mathbb{Q}) \cong \mathbb{Z}/2 \oplus \bigoplus_{p \in \mathbb{P} : p \equiv 1 (\mathrm{mod} 4)} \mathbb{Z}/p$$

Proof. The details are in Tan's article [44]. □

This shows that Mordell's theorem for elliptic curves, which says that if E is an elliptic curve, then $E(\mathbb{Q})$ is finitely generated and has finitely many torsion elements, fails for conics. In particular, the rational unit circle $C(-1)(\mathbb{Q})$ is not finitely generated, and yet consists entirely of torsion elements. Nonetheless the analogy can be salvaged by looking at rings of integers [23, p.4]. Given an algebraic number field K, write $\mathcal{O}_K$ for its ring of integers and let r_2 equal half the number of complex imbeddings of K. Shastri [40, p.68] has shown that if K is an algebraic number field in which -1 does not have a square root, then $C(-1)(\mathcal{O}_K) \cong \mathbb{Z}/4 \times \mathbb{Z}^{r_2-1}$, and that if K is an algebraic number field in which -1 does have a square root, then $C(-1)(\mathcal{O}_K) \cong \mathbb{Z}/4 \times \mathbb{Z}^{r_2}$. The interested reader is directed to Lemmermeyer [23], in which the author systematically looks at the similarities and differences between the theory of Pell conics and that of elliptic curves. In addition to the comparisons made there, the observation that there's a meaningful notion of isogeny between pointed conics, which is in particular a group homomorphism as long as the domain and codomain are non-parabolic, further strengthens the system of analogies developed by Lemmermeyer and suggests that further connections remain to be discovered.

Chapter 6

Applications to Cryptography

In this chapter we'll give a brief outline of how number-theoretic methods based on cyclic groups, elliptic curves and even conics can be used to facilitate secure communication over insecure channels.

Early cryptography presupposed that the communicating agents possessed a secret key that allowed them to encrypt their messages in a manner that (hopefully) only the other party could decrypt. An early example of this is the Caesar cipher, in which each letter is shifted a fixed number of places. For example, if that number is 2, then each copy of A is replaced by C, each copy of B is replaced by D, etc. In this way, entire messages can be translated from human-readable *plaintext* into apparently unintelligible *ciphertext* using a systematic process that the receiving party can easily reverse.

Of course, the Caesar cipher isn't too hard to break. There are only 25 possibilities, so if one knows ahead of time that a message has been encoded using a Caesar cipher, the message is readily decrypted. In modern language, there are only 25 possible keys for the Caesar cipher.

Contemporary information systems make use of much more sophisticated encryption with much larger keys. A popular algorithm is the Advanced Encryption Standard (AES). When your computer wishes to communicate with another computer securely over the internet, it is likely using AES to encrypt the data it's sending and decrypt the data it's receiving. The history is quite interesting: AES was chosen by the National Institute of Standards and Technology (NIST) after a rather lengthy competition between different teams. The AES was chosen to replace the increasingly unpopular Data Encryption Standard (DES), whose secureness had come under question due to the possibility of the existence of backdoors in the framework.

Like the Caesar cipher, AES is a symmetric cipher. What this means is that for two parties to communicate securely using AES (or a similar technology), they must both be in possession of a shared secret to use as the encryption and decryption key. Once upon time, spies would meet to exchange secret keys in order to solve this problem. But with the development of modern information technology, and in particular, the invention of public-key cryptography, this all changed.

If two parties can only communicate over an insecure channel (like the internet), it was once believed that these parties could not establish a shared secret. After all, if this secret was sent in cleartext over the channel, then it could be viewed by an eavesdropper, who would then be able to decrypt all further communications. And the secret couldn't possibly be sent in anything other than plaintext, since a shared secret had yet to be established. Thus it is easy to sympathize with the prevalence of this belief.

It was eventually realized, however, that the existence of one-way functions provides a solution. A one-way function is a function $f : X \to Y$ such that there exist currently-known efficient algorithms for computing $f(x) \in Y$ given $x \in X$, but no currently-known efficient algorithms for finding an $x \in X$ satisfying $f(x) = y$. Of course, since the definition is based on the current state of knowledge, it's difficult to give a fully mathematical account of this notion, and what once was a one-way function might no longer be cryptographically secure once better algorithms have been developed. Yet whatever the theoretical difficulties might be, one-way functions exist in practice, and thus exchanging secret keys in a public channel is possible in practice. It should be noted that although the establishment of a shared secret (called a key-exchange) is not the only application of one-way functions, it is the only one that will be considered in this thesis.

6.1 The Diffie-Hellman Key Exchange

Diffie–Hellman key exchange is a protocol that allows two parties whom have never met each other to establish a shared secret over insecure channel. It historically one of the first examples of public-key cryptography, and was invented by Ralph Merkle [26], whom decided to name it after Whitfield Diffie and Martin Hellman as a result of their contributions to the idea, especially as described in [17].

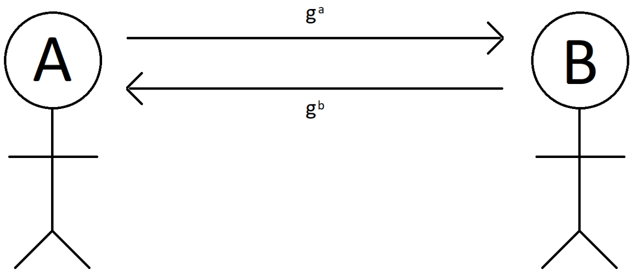
Given an abelian group G and an element $g \in G$, the function $g^{\square}$ can often be regarded as a one way function, and the classical Diffie-Hellman protocol exploits this. Not every group will work, of course. For example, if G is the set $\mathbb{R}_{>0}$ viewed as a group with respect to multiplication, then given $g \in \mathbb{R}_{>0}$, the associated inverse problem is to find $n \in \mathbb{Z}$ satisfying $g^n = h$, when both g and h are known. But this is easy! The base-g logarithm of h can be computed efficiently using standard methods, quickly revealing the integer n.

However, in the classical Diffie-Hellman protocol, the group G is taken to be the group of units of $\mathbb{Z}/p$. The inverse problem is called the discrete logarithm problem, and for a carefully chosen G tends to be very hard.

In cryptography, certain names are used with fairly consistent meaning to illustrate hypothetical agents with largely fixed abilities or goals. In particular, we'll make use of Alice and Bob, which are the standard names of two parties who wish to communicate securely. We'll also make use of Eve, a hypothetical eavesdropper with the power to read the messages that Alice and Bob send to each other, but no power to alter or modify those messages. With that in mind, the classical Diffie-Hellman protocol can be described as follows:

- Alice and Bob wish to establish a shared secret.
- Public parameters: large prime $p \in \mathbb{N}$ and non-unit $g \in (\mathbb{Z}/p)^*$.
- Alice generates a random natural number a and discloses $g^a \in (\mathbb{Z}/p)^*$.
- Bob generates a random natural number b and discloses $g^b \in (\mathbb{Z}/p)^*$.
- Bob takes the received value of g^a and raises it to the power of his secret number b, obtaining $(g^a)^b$ which equals the shared secret g^{ab}.
- Alice takes the received value of g^b and raises it to the power of her secret number a, obtaining $(g^b)^a$, which equals the shared secret g^{ab}.

Figure 6.1: The Diffie-Helman key exchange protocol



The above protocol is believed to be secure for sufficiently large primes $p \in \mathbb{N}$, as long as $p - 1$ has a very large factor. It works because an attacker (Eve) is faced with the problem of using g^a and g^b to find g^{ab}. The easiest way would be to find either a or b. For example, if Eve gains knowledge of a, she can raise the publicly disclosed value of g^b to the power of a to obtain the shared secret.

$$(g^b)^a = g^{ab}$$

More generally, if Eve can find any natural number A satisfying $g^A = g^a$, then she can raise g^b to the power of A to obtain the shared secret:

$$(g^b)^A = (g^A)^b = (g^a)^b = g^{ab}.$$

But given g^a, the problem of finding a natural number A satisfying $g^A = g^a$ is the discrete logarithm problem in $(\mathbb{Z}/p)^*$, which is potentially hard.

To ensure the discrete logarithm problem is sufficiently hard that Alice and Bob can be sure no other parties can deduce their shared secret and consequently become privy to their communications, we need to impose certain conditions on p. In particular, we require that $p - 1$ be "non-smooth", meaning that it has a very large prime factor. If this condition is not met, the Pohlig-Hellman algorithm (described in the next section) can be used to break the encryption.

6.2 Why does p have to be non-smooth?

Hereafter the discrete logarithm problem will be abbreviated DLP. The bruteforce approach to solving the DLP equation $g^n = h$ is to iteratively compute the powers

$$1, g, g^2, g^3, \ldots$$

until h is found. For appropriately chosen $g \in (\mathbb{Z}/p)^*$, this is quite a slow process, because we're forced to go through about half the elements of $(\mathbb{Z}/p)^*$ on average. Therefore the process takes $O(p)$-time on average, which is great, because p could be huge. More formally, recall that to store a number p in binary, we only need about $\hat{p} \approx \log_2(p)$-many bits. Rearranging, we find that p is approximately $2^{\hat{p}}$. Thus the bruteforce approach to the DLP takes $O(2^{\hat{p}})$ time. Since this is exponential in the amount of space $\hat{p}$ it takes to store the prime p, this plan of attack is too slow for even the most well-funded attackers.

Before proceeding, let us spend a moment dissecting the informal metaprinciple that exponential-time algorithms provide unusable attack vectors against cryptographic systems. Suppose that Alice and Bob establish a shared secret via a protocol that takes $C \cdot 2^k$ time to break, where $k \in \mathbb{N}$ is the key size (i.e. the number of bits necessary to store the shared secret) and $C \in \mathbb{R}_{>0}$ is a constant of proportionality. Then for each bit Alice and Bob include in their shared secret, the amount of resources an attacker needs to uncover their secret approximately doubles. Thus if a key size of k costs \$1 to break, then a key size of $k + 32$ costs \$$(2^{32})$ to break, which is about 4.2 billion dollars. Adding 8 bits to that puts the cost at about 1 trillion dollars, which is just under Australia's current GDP. Another 8 bits gets you to about 281 trillion dollars, which is about three times the entire Earth's current GDP. And we've only added 48 bits to our original key size! This thought experiment reassures us that the heuristic idea that exponential timescales reflect fundamentally intractable problems is reliable.

One might object that the above analysis is invalid, since it fails to account for Moore's Law, which states (in one form) that the amount computing power that a dollar can buy will roughly will double every 18 months. But even from this viewpoint, each bit added to the key size will stall potential attackers by about 18 months, and it becomes easy to stall them into failure. To make things worse for the would-be attackers, Moore's law is unrealistic over the long run, and since there are only about 2^{260} atoms in the universe, keys of size around $260 + k$ would require more dollars than are atoms in the universe. We're once again led therefore to believe in the intractability of exponential problems.

There are better methods for solving DLP than the bruteforce attack. Whereas bruteforce runs in $O(p)$-time, an algorithm called *baby-step giant-step* solves the problem in $O(\sqrt{p})$-time in the average case, a considerable improvement. However, even with this speedup, the problem remains intractably difficult. In particular, note that if $p \approx 2^{\hat{p}}$ where $\hat{p}$ is the number of bits necessary to store p, using the baby-step giant-step algorithm takes $O(\sqrt{2^{\hat{p}}})$-time, which equals $O(2^{\hat{p}/2})$. Ergo this speedup does not represent a subexponential attack and consequently does not preclude the function $g^{\square} : \mathbb{N} \to (\mathbb{Z}/p)^*$ from being used as a one-way function. It does, however, double the key sizes necessary for the same level of security.

A more severe issue occurs when $p - 1$ is a smooth integer. During the course of the development of what is now known as the Diffie-Hellman protocol, Stephen Pohlig and Martin Hellman collaborated to develop an algorithm that would break the system when $p - 1$ has many small factors [35]. Now called the Pohlig-Hellman algorithm, it has two parts.

The first part is an algorithm for solving the DLP in cyclic groups of prime-power order. The algorithm proceeds to compute each base-p digit of the desired number one at a time, using the baby-step giant-step algorithm in an appropriate quotient group to obtain the next digit. It then uses an efficient implementation of the Chinese Remainder Theorem to piece together the prime-power solutions and give the final answer. If we agree to write $u := p - 1$ in its prime-factorized form with e_i as the exponents, as in

$$u = p_1^{e_1} \dots p_n^{e_n},$$

then the worst-case time-complexity of the Pohlig-Hellman algorithm is

$$\sum_{i=1}^{n} e_i(\log n + \sqrt{p}_i).$$

Note that for smooth u, this is catastrophically subexponential. For example, suppose u equals 2^{e_1}. Then the Pohlig-Hellman algorithm takes $O(e_1 \log u)$-time in the worst case, which is $O(e_1^2)$. Since e_1 is a good approximation to $\hat{u}$, the amount of space needed to store u, this means the Pohlig-Hellman algorithm represents an $O(\hat{u}^2)$-time attack on the DLP in $\mathbb{Z}/u\mathbb{Z}$, suggesting that such choices of u are cryptographically insecure. This justifies the stipulation that $p - 1$ needs to be non-smooth for the DLP in $(\mathbb{Z}/p)^*$ to be hard.

6.3 The Diffie-Hellman key-exchange protocol over elliptic curves

Since each elliptic curve carries the structure of an abelian group in a natural way, hence elliptic curves can be used in place of the group $(\mathbb{Z}/p)^*$ in the Diffie-Hellman key-exchange protocol. Aside from a change from multiplicative to additive notation, the protocol remains largely unchanged, and can be described roughly as follows.

- Alice and Bob wish to establish a shared secret.
- Public parameters: a large finite field $\mathbb{F}_q \in \mathbb{N}$, an elliptic curve E defined over $\mathbb{F}_q$, and a non-zero element $P \in E$ that generates a sufficiently large group.
- Alice generates a random natural number a and discloses $[a]P$.
- Bob generates a random natural number b and discloses $[b]P$.
- Bob takes the received value of $[a]P$ and computes $[b]([a]P)$, which equals the shared secret $[ab]P$.

- Alice takes the received value of $[b]P$ and computes $[a]([b]P)$, which equals the shared secret $[ab]P$.

The actual representation of elliptic curves can vary between implementations. For example, Curve25519, which offers 128 bits of security, is a Montgomery curve defined over a quadratic extension of the field $\mathbb{F}_{2^{255}-19}$ and has equation

$$y^2 = x^3 + 486662x^2 + x.$$

On the other hand, NIST P-224 is in Weierstrass form, and Ed25519 is a twisted Edwards curve. A similar approach can be used to perform a Diffie-Hellman-like key exchange using pointed conics in place of elliptic curves, see Lemmermeyer [24] for example. Unfortunately, naive implementations seem to perform no better than classical Diffie-Hellman insofar as key sizes are concerned. Nonetheless, conics have been used with some level of success in cryptography; see Chen [9], for example.

6.4 The need for post-quantum cryptosystems

With the development of quantum computers, the need for quantum-resistant one-way functions and protocols is growing. For instance, cryptosystems based on the difficulty of factoring integers are broken by quantum computers armed with Shor's algorithm, which can factor an integer n of with space-complexity $u = \log_2(n)$ in $O(u^2 \log(u) \log(\log(u)))$-time when fast multiplication algorithms are used [5].

Shor has also produced algorithms for solving the discrete logarithm problem in polynomial time on a quantum computer [42]. Putting these results together, one sees that almost all of the public-key cryptosystems currently in use is hopelessly inadequate against a quantum-empowered adversary.

This prompted NIST in 2012 to launch the Post-Quantum Cryptography (PQC) project. Submissions for the first round closed in 2017, and of particular interest for our purposes is the SIDH protocol, on which one of the submissions to the PQC was based [12].

6.5 Supersingular isogeny-based Diffie-Helman (SIDH) key exchange

SIDH is a key-exchange algorithm that is believed to be quantum-resistant. The germ of the idea was described in a 1997 talk [14] entitled “Hard homogenous spaces,” and in 2006 the first cryptosystem based on this idea is published [37], which now called OIDH. In 2010, a subexponential quantum algorithm for breaking OIDH is submitted to the Arxiv [10] that exploits the commutativity of the endomorphism ring of ordinary elliptic curves. Just a year later, a supersingular variant is published [22] that thwarts the aforementioned attack as a consequence non-commutativity of the endomorphism ring in the supersingular case. It is this variant, termed SIDH, that we will concern ourselves with here.

Although the basic idea behind SIDH can be explained in elementary terms, nonetheless the technical details are fairly demanding. Consequently this will be a somewhat brief explanation of the protocol. The difference between elliptic curve Diffie-Hellman and SIDH can be summarized by contrasting the hard problems on which they're based. In elliptic curve Diffie-Hellman, Eve is privy to the public point P on the curve and Alice's disclosure of $[a]P$, and is faced with the task of using this information to discover a. In SIDH, on the other hand, Eve is privy to a public supersingular curve E and to its image $q_A(E)$ under an isogeny q_A, and is faced with the task of finding q_A. The commutativity relation $[a]([b]P) = [b]([a]P)$ which in elliptic curve Diffie-Hellman ensures that both Alice and Bob arrive at the same shared secret is replaced by the commutativity of the diagram

$$\begin{tikzcd}
E \arrow[r, "q_B"] \arrow[d, "q_A"'] & E/B \arrow[d, "q'_A"] \\
E/A \arrow[r, "q'_B"] & E/(A+B).
\end{tikzcd}$$

With that in mind, the protocol can be described (roughly) as follows:

- Alice and Bob wish to establish a shared secret.
- Public parameters:
  - natural numbers e_A and e_B whose size is related to the key size
  - a natural number f with the property that the quantity $p := f \cdot 2^{e_A} 3^{e_B} - 1$ is prime.

  - an elliptic curve E defined over $\mathbb{F}_{p^2}$ with $(p+1)^2$ elements
  - a two-element generating set $\{P_A, Q_A\}$ for the group $E[2^{e_A}]$ of all $P \in E$ such that $[2^{e_A}]P = 0$
  - a two-element generating set $\{P_B, Q_B\}$ for the group $E[2^{e_B}]$ of all $P \in E$ such that $[2^{e_B}]P = 0$

- Alice generates a pair of random natural numbers (m_A, n_A) that she keeps secret, and computes $R_A := [m_A]P_A + [n_A]Q_A$. She computes the isogeny $\varphi_A : E \to E/R_A$ and discloses the coefficients of the curve E/R_A while keeping φ_A secret.

- Bob generates a pair of random natural numbers (m_B, n_B) that she keeps secret, and computes $R_B := [m_B]P_B + [n_B]Q_B$. He discloses the coefficients of the curve E/R_B. He computes the isogeny $\varphi_B : E \to E/R_B$ and discloses the coefficients of the curve E/R_B while keeping φ_B secret.

- Bob helps Alice out by also disclosing $\varphi_B(P_A)$ and $\varphi_B(Q_A)$.

- Alice helps Bob out by disclosing $\varphi_A(P_B)$ and $\varphi_B(Q_A)$.

- Alice takes the curve E/R_B that she received from Bob, and using the points $\varphi_B(P_A)$ and $\varphi_B(Q_A)$ on this curve which she also received, computes the linear combination $\varphi_B(R_A) := [m_A]\varphi_B(P_A) + [n_A]\varphi_B(Q_A)$. She uses this to build the quotient
$$(E/R_B)/(R'_A).$$
The j-invariant of this curve is the shared-secret.

- Bob takes the curve E/R_A that he received from Alice, and using the points $\varphi_A(P_B)$ and $\varphi_A(Q_B)$ on this curve which he also received, computes the linear combination $\varphi_A(R_B) := [m_B]\varphi_B(P_B) + [n_B]\varphi_A(Q_B)$. He uses this to build the quotient
$$(E/R_B)/(R'_A).$$
The j-invariant of this curve is the shared-secret.

The SIDH key exchange protocol has a major advantage over other post-quantum key exchange protocols insofar as the key sizes are smaller for the same security level; however, the computational time needed to execute the protocol is much higher than for the major competing protocols [13]. This does not matter too much on the client-side, but for a server that needs to service hundreds of requests each second, the current slowness of SIDH implementations makes it a less attractive contender. In light of the existence of a meaningful notion of isogenies between conics, and the comparative simplicity of conic arithmetic, it seems reasonable to investigate the possibility of using 'higher-dimensional conics' in place of elliptic curves in an SIDH-like scheme. If it's possible to find higher-dimensional degree 2 varieties with useful group structures and a good theory of isogenies and quotients, they could offer a means of improving the efficiency of the protocol.

Appendices

Appendix A

Python code for Velu's Formula

```
#### Parameters

# Define the elliptic curve by its "a" coefficients

a = dict()

a[1] = 1
a[3] = 1
a[2] = -1
a[4] = -3
a[6] = 3

# A special subset S of the subgroup of interest that contains half the elements that
s = [[1,-2],[-1,-2],[3,-6]]

#### Code

# Define a function w whose vanishing set is the curve. This is not used in the formul

def w(P):
x = P[0]
y = P[1]
return (y**2 + a[1] * x*y+a[3]*y)-(x**3 + a[2]*x**2 + a[4]*x + a[6])
```

```
# Compute the corresponding "b" coefficients

b = dict()

b[2] = a[1]**2 + 4*a[2]
b[4] = a[1]*a[3] + 2*a[4]
b[6] = a[3]**2+4*a[6]
b[8] = (a[1]**2)*a[6] - a[1]*a[3]*a[4]+4*a[2]*a[6]+a[2]*(a[3]**3)-a[4]**2

# Define the two partial derivatives of the defining function of the curve, as well as

def wx(P):
x = P[0]
y = P[1]
return -3*(x**2) - 2*a[2]*x - a[4] + a[1]*y

def wy(P):
x = P[0]
y = P[1]
return 2*y + a[1]*x + a[3]

def t(P):
x = P[0]
y = P[1]
return 6*(x**2) + b[2]*x + b[4]

def u(P):
return wy(P)**2

# Compute the constants T and W

T = sum([t(P) for P in s])
W = sum([u(P)+P[0]*t(P) for P in s])

# Compute the coefficients of the quotient curve
```

```
A = [None for _ in range(1,8)]
A[1] = a[1]
A[3] = a[3]
A[2] = a[2]
A[4] = a[4]-5*T
A[6] = a[6]-b[2]*T-7*W

# Print the relevant information

print("wx:",[wx(P) for P in s])
print("wy:",[wy(P) for P in s])
print("t:",[t(P) for P in s])
print("u:",[u(P) for P in s])
print("T:",T,"W:",W)
print("A[1,3,2,4,6] = ",end = "")
print(A[1],A[3],A[2],A[4],A[6], sep = ", ")
```

Bibliography

[1] Avner Ash and Robert Gross. *Elliptic tales: curves, counting, and number theory.* Princeton University Press, 2014.

[2] Robert B Ash. A course in commutative algebra, chapter 5. `https://faculty.math.illinois.edu/~r-ash/ComAlg/ComAlg5.pdf`.

[3] Arthur OL Atkin. The number of points on an elliptic curve modulo a prime. *preprint*, 1988.

[4] azimut (https://math.stackexchange.com/users/61691/azimut). Why is the localization at a prime ideal a local ring? Mathematics Stack Exchange. `https://math.stackexchange.com/q/300459(version:2013-02-11)`.

[5] David Beckman, Amalavoyal N Chari, Srikrishna Devabhaktuni, and John Preskill. Efficient networks for quantum factoring. *Physical Review A*, 54(2):1034, 1996.

[6] Daniel J Bernstein, Peter Birkner, Marc Joye, Tanja Lange, and Christiane Peters. Twisted edwards curves. In *International Conference on Cryptology in Africa*, pages 389–405. Springer, 2008.

[7] Manjul Bhargava and Arul Shankar. Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves. *Annals of Mathematics*, pages 191–242, 2015.

[8] Andries E Brouwer. Regular differential forms. `https://www.win.tue.nl/~aeb/2WF02/difffm.pdf`.

[9] Zhi-Gang Chen and Xin-Xia Song. A public-key cryptosystem scheme on conic curves over z n. In *Machine Learning and Cybernetics, 2007 International Conference on*, volume 4, pages 2183–2187. IEEE, 2007.

[10] A. M. Childs, D. Jao, and V. Soukharev. Constructing elliptic curve isogenies in quantum subexponential time. *ArXiv e-prints*, December 2010.

[11] Pete L Clark. Commutative algebra, 2015. `http://math.uga.edu/~pete/integral2015.pdf`.

[12] Sike Contributors. Sike resources page. `https://sike.org/#resources`.

[13] Craig Costello, Patrick Longa, and Michael Naehrig. Efficient algorithms for supersingular isogeny diffie-hellman. In *Annual Cryptology Conference*, pages 572–601. Springer, 2016.

[14] Jean Marc Couveignes. Hard homogeneous spaces. *IACR Cryptology ePrint Archive*, 2006:291, 2006.

[15] Brian A Davey and Hilary A Priestley. *Introduction to lattices and order.* Cambridge university press, 2002.

[16] Max Deuring. Die typen der multiplikatorenringe elliptischer funktionenkörper. In *Abhandlungen aus dem mathematischen Seminar der Universität Hamburg*, volume 14, pages 197–272. Springer, 1941.

[17] Whitfield Diffie and Martin Hellman. New directions in cryptography. *IEEE transactions on Information Theory*, 22(6):644–654, 1976.

[18] Marcel Erné, Jürgen Koslowski, Austin Melton, and George E Strecker. A primer on galois connections. *Annals of the New York Academy of Sciences*, 704(1):103–125, 1993.

[19] Tim Güneysu, Christof Paar, and Jan Pelzl. On the security of elliptic curve cryptosystems against attacks with special-purpose hardware. *Special-Purpose Hardware for Attacking Cryptographic Systems–SHARCS'06*, pages 03–04, 2006.

[20] Robin Hartshorne. *Algebraic geometry*, volume 52. Springer Science & Business Media, 2013.

[21] Mel Hochester. Dimension theory and systems of parameters. `http://www.math.lsa.umich.edu/~hochster/615W10/supDim.pdf`.

[22] David Jao and Luca De Feo. Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies. In *International Workshop on Post-Quantum Cryptography*, pages 19–34. Springer, 2011.

[23] Franz Lemmermeyer. Conics-a poor man's elliptic curves. *arXiv preprint math/0311306*, 2003.

[24] Franz Lemmermeyer. Introduction to cryptography, 2006.

[25] Barry Mazur. Rational points on modular curves. In *Modular functions of one variable V*, pages 107–148. Springer, 1977.

[26] Ralph C Merkle. Secure communications over insecure channels. *Communications of the ACM*, 21(4):294–299, 1978.

[27] James S Milne. *Algebraic geometry*. Allied Publishers, 1996.

[28] Steve Mitchell. Supplementary field theory notes. `https://sites.math.washington.edu/~mitchell/Algg/field.pdf`.

[29] Peter L Montgomery. Speeding the pollard and elliptic curve methods of factorization. *Mathematics of computation*, 48(177):243–264, 1987.

[30] LJ Mordell. On the integer solutions of the equation ey2= ax3+ bx2+ cx+ d. *Proceedings of the London Mathematical Society*, 2(1):415–419, 1923.

[31] David Mumford, John Fogarty, and Frances Kirwan. *Geometric invariant theory*, volume 34. Springer Science & Business Media, 1994.

[32] David Mumford, Chidambaran Padmanabhan Ramanujam, and Jurij Ivanovič Manin. *Abelian varieties*, volume 108. Oxford university press Oxford, 1974.

[33] Katsuyuki Okeya, Hiroyuki Kurumatani, and Kouichi Sakurai. Elliptic curves with the montgomery-form and their cryptographic applications. In Hideki Imai and Yuliang Zheng, editors, *Public Key Cryptography*, pages 238–257, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.

[34] Alan R Pears. *Dimension theory of general spaces*, volume 222. Cambridge University Press London, 1975.

[35] Stephen Pohlig and Martin Hellman. An improved algorithm for computing logarithms over gf (p) and its cryptographic significance (corresp.). *IEEE Transactions on information Theory*, 24(1):106–110, 1978.

[36] Bjorn Poonen. Lectures on rational points on curves, 2006. `https://math.mit.edu/~poonen/papers/curves.pdf`.

[37] Alexander Rostovtsev and Anton Stolbunov. Public-key cryptosystem based on isogenies. *IACR Cryptology ePrint Archive*, 2006:145, 2006.

[38] Rene Schoof. Elliptic curves over finite fields and the computation of square roots mod p. *Mathematics of Computation*, 44(170):483, 1985.

[39] René Schoof. Counting points on elliptic curves over finite fields. *J. Théor. Nombres Bordeaux*, 7(1):219–254, 1995.

[40] Parvati Shastri. Integral points on the unit circle. *Journal of Number Theory*, 91(1):67–70, 2001.

[41] Shailesh A Shirali. Groups associated with conics. *The Mathematical Gazette*, 93(526):27–41, 2009.

[42] Peter W Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, 41(2):303–332, 1999.

[43] Joseph H. Silverman. *The arithmetic of elliptic curves*. World Publishing Company, 2011.

[44] Lin Tan. The group of rational points on the unit circle. *Mathematics Magazine*, 69(3):163–171, 1996.

[45] Jacques Vélu. Isogénies entre courbes elliptiques. *Comptes Rendus de l'Academie des Sciences de Paris*, 273, 1971.

[46] Alex Wright. Transcendence degree. `http://www-personal.umich.edu/~alexmw/TranscDeg.pdf`.