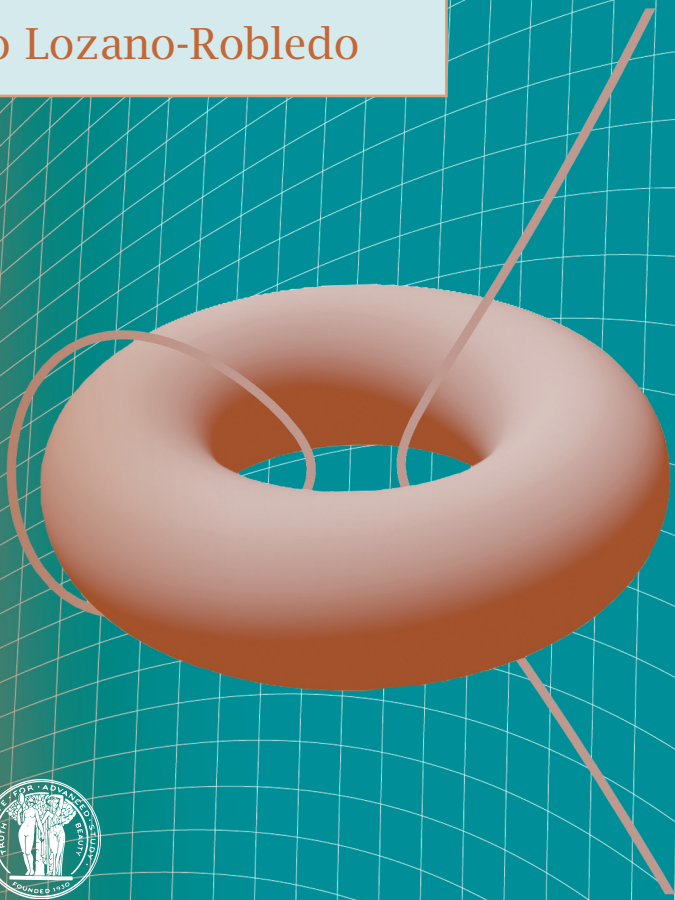
STUDENT MATHEMATICAL LIBRARY
IAS/PARK CITY MATHEMATICAL SUBSERIES
Volume 58

Elliptic Curves, Modular Forms, and Their L-functions

Álvaro Lozano-Robledo



American Mathematical Society
Institute for Advanced Study

Elliptic Curves, Modular Forms, and Their L-functions

STUDENT MATHEMATICAL LIBRARY
IAS/PARK CITY MATHEMATICAL SUBSERIES
Volume 58

Elliptic Curves, Modular Forms, and Their L-functions

Álvaro Lozano-Robledo


American Mathematical Society, Providence, Rhode Island
Institute for Advanced Study, Princeton, New Jersey

Cover art courtesy of Karl Rubin, using MegaPOV, which is based on POV-Ray, both of which are open source, freely available software.

2000 *Mathematics Subject Classification.* Primary 14H52, 11G05; Secondary 11F03, 11G40.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-58

Library of Congress Cataloging-in-Publication Data

Lozano-Robledo, Álvaro, 1978–
Elliptic curves, modular forms, and their L-functions / Álvaro Lozano-Robledo.
p. cm. — (Student mathematical library ; v. 58. IAS/Park City mathematical subseries)
Includes bibliographical references and index.
ISBN 978-0-8218-5242-2 (alk. paper)
1. Curves, Elliptic. 2. Forms, Modular. 3. L-functions. 4. Number theory. I. Title.
QA567.2.E44L69 2010
516.3′52—dc22

2010038952

10 9 8 7 6 5 4 3 2 1 16 15 14 13 12 11

A mis padres, que me enseñaron todo lo importante,
a mi abuela, por su sonrisa que no aparece en fotografías,
a Marisa, por lograr que siempre me supere,
y a "Sally", que nacerá pronto.

Contents

Preface

This book grew out of the lecture notes for a course on "Elliptic Curves, Modular Forms and L-functions" that the author taught at an undergraduate summer school as part of the 2009 Park City Mathematics Institute. These notes are an *introductory survey* of the theory of elliptic curves, modular forms and their L-functions, with an emphasis on examples rather than proofs. The main goal is to provide the reader with a *big picture* of the surprising connections among these three types of mathematical objects, which are seemingly so distinct. In that vein, one of the themes of the book is to explain the statement of the modularity theorem (Theorem 5.4.6), previously known as the Taniyama-Shimura-Weil conjecture (Conjecture 5.4.5). In order to underscore the importance of the modularity theorem, we also discuss in some detail one of its most renowned consequences: Fermat's last theorem (Example 1.1.5 and Section 5.5).

It would be impossible to give the proofs of the main theorems on elliptic curves and modular forms in one single course, and the proofs would be outside the scope of the undergraduate curriculum. However, the definitions, the statements of the main theorems and their corollaries can be easily understood by students with some standard undergraduate background (calculus, linear algebra, elementary number theory and a first course in abstract algebra). Proofs that are accessible to a student are left to the reader and proposed as exercises

at the end of each chapter. The reader should be warned, though, that there are multiple references to mathematical objects and results that we will not have enough space to discuss in full, and the student will have to take these items on faith (we will provide references to other texts, however, for those students who wish to deepen their understanding). Some other objects and theorems are mentioned in previous chapters but only explained fully in later chapters. To avoid any confusion, we always try to clarify in the text which objects or results the student should take on faith, which ones we expect the student to be familiar with, and which will be explained in later chapters (by providing references to later sections of the book).

The book begins with some motivating problems, such as the congruent number problem, Fermat's last theorem, and the representations of integers as sums of squares. Chapter 2 is a survey of the algebraic theory of elliptic curves. In Section 2.9, we give a proof of the weak Mordell-Weil theorem for elliptic curves with rational 2-torsion and explain the method of 2-descent. The goal of Chapter 3 is to motivate the connection between elliptic curves and modular forms. To that end, we discuss complex lattices, tori, modular curves and how these objects relate to elliptic curves over the complex numbers. Chapter 4 introduces the spaces of modular forms for $\mathrm{SL}(2,\mathbb{Z})$ and other congruence subgroups (e.g., $\Gamma_0(N)$). In Chapter 5 we define the L-functions attached to elliptic curves and modular forms. We briefly discuss the Birch and Swinnerton-Dyer conjecture and other related conjectures. Finally, in Section 5.4, we justify the statement of the original conjecture of Taniyama-Shimura-Weil (which we usually refer to as the modularity theorem, since it was proved in 1999); i.e., we explain the surprising connection between elliptic curves and certain modular forms, and justify which modular forms correspond to elliptic curves.

In order to make this book as self-contained as possible, I have also included five appendices with concise introductions to topics that some students may not have encountered in their classes yet. Appendix A is a quick reference guide to two popular software packages: PARI and Sage. Throughout the book, we strongly recommend that the reader tries to find examples and do calculations using one of these

two packages. Appendix B is a brief summary of complex analysis. Due to space limitations we only include definitions, a few examples, and a list of the main theorems in complex analysis; for a full treatment see [**Ahl79**], for instance. In Appendix C we introduce the projective line and the projective plane. The p-adic integers and the p-adic numbers are treated in Appendix D (for a complete reference, see [**Gou97**]). Finally, in Appendix E we list infinite families of elliptic curves over $\mathbb{Q}$, one family for each of the possible torsion subgroups over $\mathbb{Q}$.

I would like to emphasize once again that this book is, by no means, a thorough treatment of elliptic curves and modular forms. The theory is far too vast to be covered in one single volume, and the proofs are far too technical for an undergraduate student. Therefore, the humble goals of this text are to provide a *big picture* of the vast and fast-growing theory, and to be an "advertisement" for undergraduates of these very active and exciting areas of number theory. The author's only hope is that, after reading this text, students will feel compelled to study elliptic curves and modular forms in depth, and in all their full glory.

There are many excellent references that I would recommend to the students, and that I have frequently consulted in the preparation of this book:

(1) There are not that many books on these subjects at the *undergraduate level*. However, Silverman and Tate's book [**SiT92**] is an excellent introduction to elliptic curves for undergraduates. Washington's book [**Was08**] is also accessible for undergraduates and emphasizes the cryptography applications of elliptic curves. Stein's book [**Ste08**] also has an interesting chapter on elliptic curves.

(2) There are several *graduate-level* texts on elliptic curves. Silverman's book [**Sil86**] is the standard reference, but Milne's [**Mil06**] is also an excellent introduction to the theory of elliptic curves (and also includes a chapter on modular forms). Before reading Silverman or Milne, the reader would benefit

from studying some algebraic geometry and algebraic number theory. (Milne's book does not require as much algebraic geometry as Silverman's.)

(3) The theory of modular forms and L-functions is definitely a *graduate topic*, and the reader will need a strong background in algebra to understand all the fine details. Diamond and Shurman's book [**DS05**] contains a neat, modern and thorough account of the theory of modular forms (including much information about the modularity theorem). Koblitz's book [**Kob93**] is also a very nice introduction to the theory of elliptic curves and modular forms (and includes a lot of information about the congruent number problem). Chapter 5 in Milne's book [**Mil06**] contains a good, concise overview of the subject. Serre's little book [**Ser77**] is always worth reading and also contains an introduction to modular forms. Miyake's book [**Miy06**] is a very useful reference.

(4) Finally, if the reader is interested in computations, we recommend Cremona's [**Cre97**] or Stein's [**Ste07**] book. If the reader wants to play with fundamental domains of modular curves, try Helena Verrill's applet [**Ver05**].

I would like to thank the organizers of the undergraduate summer school at PCMI, Aaron Bertram and Andrew Bernoff, for giving me the opportunity to lecture in such an exciting program. Also, I would like to thank Ander Steele and Aaron Wood for numerous corrections and comments of an early draft. Last, but not least, I would like to express my gratitude to Keith Conrad, David Pollack and William Stein, whose abundant comments and suggestions have improved this manuscript much more than it would be safe to admit.

Álvaro Lozano-Robledo

Chapter 1

Introduction

Notation:

$\mathbb{N} = \{1, 2, 3, \ldots\}$ is the set of natural numbers.

$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ is the ring of integers.

$\mathbb{Q} = \{\frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0\}$ is the field of rational numbers.

$\mathbb{R}$ is the field of real numbers.

$\mathbb{C} = \{a + bi : a, b \in \mathbb{R},\ i^2 = -1\}$ is the field of complex numbers.

In this chapter, we introduce elliptic curves, modular forms and L-functions through examples that motivate the definitions.

1.1. Elliptic curves

For the time being, we define an elliptic curve to be any equation of the form

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

with $a, b, c \in \mathbb{Z}$ and such that the polynomial $x^3 + ax^2 + bx + c$ does not have repeated roots. See Section 2.2 for a precise definition.

Example 1.1.1. *Are there three consecutive integers whose product is a perfect square?*

There are some trivial examples that involve the number zero, for example, $0, 1$ and 2, whose product equals $0 \cdot 1 \cdot 2 = 0 = 0^2$, a square.

Are there any non-trivial examples? If we try to assign variables to our problem, we see that we are trying to find solutions to

$$y^2 = x(x+1)(x+2) \tag{1.1}$$

with $x, y \in \mathbb{Z}$ and $y \neq 0$. Equation (1.1) defines an elliptic curve. It turns out that there are no integral solutions other than the trivial ones (see Exercise 1.4.1). Are there rational solutions, i.e., are there solutions with $x, y \in \mathbb{Q}$? This is a more delicate question, but the answer is still no (we will prove it in Example 2.7.6). Here is a similar question, with a very different answer:

- *Are there three integers that differ by* 5*, i.e.,* x*,* $x+5$ *and* $x+10$*, and whose product is a perfect square?*

In this case, we are trying to find solutions to $y^2 = x(x+5)(x+10)$ with $x, y \in \mathbb{Z}$. As in the previous example, there are trivial solutions (those which involve 0) but in this case, there are non-trivial solutions as well:

$$\begin{aligned} (-9)\cdot(-9+5)\cdot(-9+10) = (-9)\cdot(-4)\cdot 1 = 36 &= 6^2 \\ 40\cdot(40+5)\cdot(40+10) = 40\cdot 45\cdot 50 = 90000 &= 300^2. \end{aligned}$$

Moreover, there are also *rational* solutions, which are far from obvious:

$$\begin{aligned} \left(\frac{5}{4}\right)\cdot\left(\frac{5}{4}+5\right)\cdot\left(\frac{5}{4}+10\right) &= \left(\frac{75}{8}\right)^2 \\ \left(-\frac{50}{9}\right)\cdot\left(-\frac{50}{9}+5\right)\cdot\left(-\frac{50}{9}+10\right) &= \left(\frac{100}{27}\right)^2 \end{aligned}$$

and, in fact, there are infinitely many *rational* solutions! Here are some of the x-coordinates that work:

$$x = -9,\ 40,\ \frac{5}{4},\ \frac{-50}{9},\ \frac{961}{144},\ \frac{7200}{961},\ -\frac{12005}{1681},\ -\frac{16810}{2401},\ -\frac{27910089}{5094049}, \ldots$$

In Sections 2.9 and 2.10 we will explain a method to find rational points on elliptic curves and, in Exercise 2.12.23, the reader will calculate all the rational points of $y^2 = x(x+5)(x+10)$. ■

Example 1.1.2 (The Congruent Number Problem)**.** *We say that* $n \geq 1$ *is a congruent number if there exists a right triangle whose sides are rational numbers and whose area equals* n*. What natural numbers are congruent?*

For instance, the number 6 is congruent, because the right triangle with sides of length $(a, b, c) = (3, 4, 5)$ has area equal to $\frac{3\cdot 4}{2} = 6$. Similarly, the number 30 is the area of the right triangle with sides $(5, 12, 13)$; thus, 30 is a congruent number.

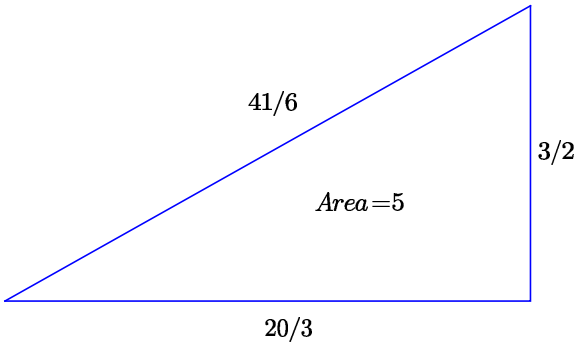


Figure 1. A right triangle of area 5 and rational sides.

The number 5 is congruent but there is no right triangle with integer sides and area equal to 5. However, our definition allowed *rational* sides, and the triangle with sides $\left(\frac{3}{2}, \frac{20}{3}, \frac{41}{6}\right)$ has area exactly 5. We do not allow, however, triangles with irrational sides even if the area is an integer. For example, the right triangle $(1, 2, \sqrt{5})$ has area 1, but that does not imply that 1 is a congruent number (in fact, 1 is *not* a congruent number, as we shall see below).

The congruent number problem is one of the oldest open problems in number theory. For more than a millennium, mathematicians have attempted to provide a characterization of all congruent numbers. The oldest written record of the problem dates back to the early Middle Ages, when it appeared in an Arab manuscript written before 972 (a later 10th century manuscript written by Mohammed Ben Alcohain would go as far as to claim that the principal object of the theory of rational right triangles is to find congruent numbers). It is known that Leonardo Pisano, a.k.a. *Fibonacci*, was challenged around 1220 by Johannes of Palermo to find a rational right triangle of area

$n = 5$, and Fibonacci found the triangle $(\frac{3}{2}, \frac{20}{3}, \frac{41}{6})$. We will explain a method to find this triangle below. In 1225, Fibonacci wrote a more general treatment about the congruent number problem, in which he stated (without proof) that if n is a perfect square, then n cannot be a congruent number. The proof of such a claim had to wait until Pierre de Fermat (1601-1665) settled that the number 1 (and every square number) is not a congruent number (a result that he showed in order to prove the case $n = 4$ of Fermat's last theorem).

The connection between the congruent number problem and elliptic curves is as follows:

Proposition 1.1.3. *The number $n > 0$ is congruent if and only if the curve $y^2 = x^3 - n^2x$ has a point (x, y) with $x, y \in \mathbb{Q}$ and $y \neq 0$. More precisely, there is a one-to-one correspondence $C_n \longleftrightarrow E_n$ between the following two sets:*

$$\begin{aligned} C_n &= \{(a, b, c) : a^2 + b^2 = c^2,\ \frac{ab}{2} = n\} \\ E_n &= \{(x, y) : y^2 = x^3 - n^2x,\ y \neq 0\}. \end{aligned}$$

Mutually inverse correspondences $f : C_n \to E_n$ and $g : E_n \to C_n$ are given by

$$f((a, b, c)) = \left(\frac{nb}{c - a}, \frac{2n^2}{c - a}\right),\ g((x, y)) = \left(\frac{x^2 - n^2}{y}, \frac{2nx}{y}, \frac{x^2 + n^2}{y}\right).$$

The reader can provide a proof (see Exercise 1.4.3). For example, the curve $E : y^2 = x^3 - 25x$ has a point $(-4, 6)$ that corresponds to the triangle $(\frac{3}{2}, \frac{20}{3}, \frac{41}{6})$. But E has other points, such as $(\frac{1681}{144}, \frac{62279}{1728})$ that corresponds to the triangle

$$\left(\frac{1519}{492}, \frac{4920}{1519}, \frac{3344161}{747348}\right)$$

which also has area equal to 5. See Figure 2.

Today, there are partial results toward the solution of the congruent number problem, and strong results that rely heavily on famous (and widely accepted) conjectures, but we do not have a full answer yet. For instance, in 1975 (see **[Ste75]**), Stephens showed that the Birch and Swinnerton-Dyer conjecture (which we will discuss in Section 5.2) implies that any positive integer $n \equiv 5$, 6 or 7 mod 8 is a

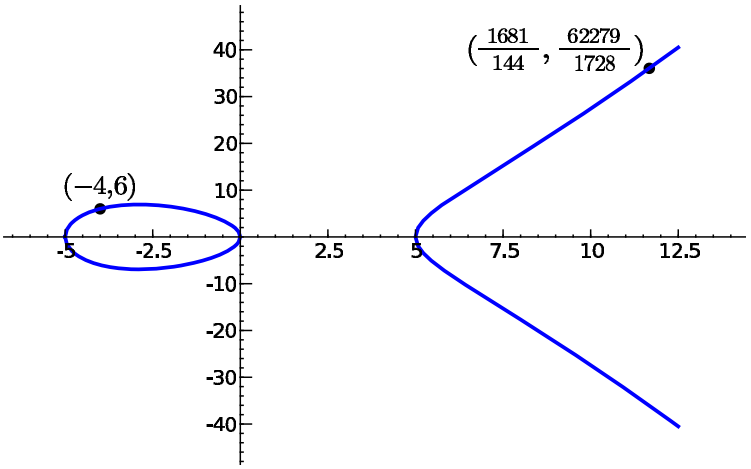


Figure 2. Two rational points on the curve $y^2 = x^3 - 25x$.

congruent number. For example, $n = 157 \equiv 5 \mod 8$ must be a congruent number and, indeed, Don Zagier has exhibited a right triangle (a, b, c) whose area equals 157. The hypotenuse of the simplest such triangle is:

$$c = \frac{224403517704336969924557513090667486316094847204 1}{8912332268928859588025535178967163570016480830}.$$

In Example 5.2.7 we will see an application of the conjecture of Birch and Swinnerton-Dyer to find a rational point P on $y^2 = x^3 - 157^2 x$, which corresponds to a right triangle of area 157 via the correspondence in Proposition 1.1.3.

The best known result on the congruent number problem is due to J. Tunnell:

Theorem 1.1.4 (Tunnell, 1983, [**Tun83**]). *If n is an odd square-free positive integer and n is the area of a right triangle with rational sides, then the following cardinalities are equal:*

$$\begin{aligned} &\#\{(x, y, z) \in \mathbb{Z}^3 : n = 2x^2 + y^2 + 32z^2\} \\ &= \frac{1}{2}\left(\#\left\{(x, y, z) \in \mathbb{Z}^3 : n = 2x^2 + y^2 + 8z^2\right\}\right) \end{aligned}$$

and, if n is even,

$$\#\{(x,y,z)\in\mathbb{Z}^3 : \frac{n}{2} = 4x^2+y^2+32z^2\}$$
$$= \frac{1}{2}\left(\#\{(x,y,z)\in\mathbb{Z}^3 : \frac{n}{2} = 4x^2+y^2+8z^2\}\right).$$

Moreover, if the Birch and Swinnerton-Dyer conjecture is true, then, conversely, these equalities imply that n is a congruent number.

For example, for $n=2$ we have $\frac{n}{2} = 1 = 4x^2+y^2+32z^2$ if and only if $x=z=0$ and $y=\pm 1$, so the left-hand side of the appropriate equation in Tunnell's theorem is equal to 2. However, the right-hand side is equal to 1 and the equality does not hold. Hence, 2 is *not* a congruent number.

For a complete historical overview of the congruent number problem, see **[Dic05]**, Ch. XVI. The book **[Kob93]** contains a thorough modern treatment of the problem. The reader may also find useful an expository paper **[Con08]** on the congruent number problem, written by Keith Conrad. Another neat exposition, more computational in nature (using Sage), appears in **[Ste08]**, Section 6.5.3. ■

Example 1.1.5 (Fermat's last theorem)**.** *Let $n \geq 3$. Are there any solutions to $x^n + y^n = z^n$ in integers x, y, z with $xyz \neq 0$?* The answer is *no*. In 1637, Pierre de Fermat wrote in the margin of a book (Diophantus' *Arithmetica*; see Figure 9 in Section 5.5) that he had found a marvellous proof, but the margin was too small to contain it. Since then, many mathematicians tried in vain to demonstrate (or disprove!) this claim. A proof was finally found in 1995 by Andrew Wiles (**[Wil95]**). We shall discuss the proof in some more detail in Section 5.5. For now, we will outline the basic structure of the argument.

First, it is easy to show that, to prove the theorem, it suffices to show the cases $n=4$ and $n=p\geq 3$, a prime. It is not difficult to show that $x^4+y^4=z^4$ has no non-trivial solutions in $\mathbb{Z}$ (this was first shown by Fermat). Now, suppose that $p\geq 3$ and a,b,c are integers with $abc\neq 0$ and $a^p+b^p=c^p$. Gerhard Frey conjectured that if such a triple of integers exists, then the elliptic curve

$$E : y^2 = x(x-a^p)(x+b^p)$$

Figure 3. Pierre de Fermat (1601-1665).

would have some unexpected properties that would contradict a well-known conjecture that Taniyama, Shimura and Weil had formulated in the 1950's. Their conjecture spelled out a strong connection between elliptic curves and modular forms, which we will describe in Section 5.4. Ken Ribet proved that, indeed, such a curve would contradict the Taniyama-Shimura-Weil (TSW) conjecture. Finally, Andrew Wiles was able to prove the TSW conjecture in a special case that would cover the hypothetical curve E. Therefore, E cannot exist and the triple (a, b, c) cannot exist, either.

The Taniyama-Shimura-Weil conjecture (Conjecture 5.4.5), i.e., the modularity theorem 5.4.6, was fully proved by Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor in their article [**BCDT01**]. ■

1.2. Modular forms

Let $\mathbb{C}$ be the complex plane and let $\mathbb{H}$ be the upper half of the complex plane, i.e., $\mathbb{H} = \{a+bi : a, b \in \mathbb{R}, b > 0\}$. A *modular form* is a function

$f : \mathbb{H} \to \mathbb{C}$ that has several relations among its values (which we will specify in Definitions 4.1.3 and 4.2.1). In particular, the values of the function f satisfy several types of periodicity relations. For example, the modular forms for $\mathrm{SL}(2,\mathbb{Z})$ satisfy, among other properties, the following:

- $f(z) = f(z+1)$ for all $z \in \mathbb{H}$, and
- $f\left(\frac{-1}{z}\right) = z^k f(z)$ for all $z \in \mathbb{H}$. The number k is an integer called the *weight* of the modular form.

We will describe modular forms in detail in Chapter 4. Let us see some examples that motivate our interest in these functions.

Example 1.2.1 (Representations of integers as sums of squares)**.** *Is the number $n > 0$ a sum of two (integer) squares?* In other words, are there $a, b \in \mathbb{Z}$ such that $n = a^2 + b^2$? And if so, in how many different ways can you represent n as a sum of two squares?

For instance, the number $n = 3$ cannot be represented as a sum of two squares but the number $n = 5$ has 8 distinct representations:

$$5 = (\pm 1)^2 + (\pm 2)^2 = (\pm 2)^2 + (\pm 1)^2.$$

Notice that here we consider $(-1)^2 + 2^2$, $1^2 + 2^2$ and $2^2 + 1$ as *distinct* representations of 5. A general formula for the number of representations of an integer n as a sum of 2 squares, due to Lagrange, Gauss and Jacobi, is given by

$$S_2(n) = 2\left(1 + \left(\frac{-1}{n}\right)\right) \sum_{d|n} \left(\frac{-1}{d}\right), \tag{1.2}$$

where $\left(\frac{m}{n}\right)$ is the Jacobi symbol and $\sum_{d|n}$ is a sum over all positive divisors of n (including 1 and n). Here we just need the easiest values $\left(\frac{-1}{n}\right) = (-1)^{(n-1)/2}$ of the Jacobi symbol. Let us see that the formula works:

$$S_2(3) = 2\left(1 + \left(\frac{-1}{3}\right)\right) \sum_{d|3} \left(\frac{-1}{d}\right) = 2(1 + (-1))(1 + (-1)) = 0,$$

$$S_2(5) = 2\left(1 + \left(\frac{-1}{5}\right)\right) \sum_{d|5} \left(\frac{-1}{d}\right) = 2(1 + 1)(1 + 1) = 8,$$

and $S_2(9) = 4$. Indeed, the number nine has 4 different representations: $9 = (\pm 3)^2 + 0^2 = 0^2 + (\pm 3)^2$. Let us explore other similar questions.

Let $n > 0$ and $k \geq 2$. Is the number $n > 0$ a sum of k (integer) squares? In other words, are there $a_1, \ldots, a_k \in \mathbb{Z}$ such that $n = a_1^2 + \cdots + a_k^2$? And if so, in how many different ways can you represent n as a sum of k squares? Lagrange showed that *every* natural number can be represented as a sum of $k \geq 4$ squares, but how many different representations are there?

Let $S_k(n)$ be the number of representations of n as a sum of k squares. Determining exact formulas for $S_k(n)$ is a classical problem in number theory. There are exact formulas known in a number of cases (e.g. Eq. 1.2). The formulas for $k = 4, 6$ and 8 are due to Jacobi and Siegel. We write $n = 2^\nu g$, with $\nu \geq 0$ and odd $g > 0$:

$$\begin{aligned} S_4(n) &= 8 \sum_{d|n,\ 4\nmid d} d, \\ S_6(n) &= \left(\left(\frac{-1}{g} \right) 2^{2\nu+4} - 4 \right) \sum_{d|g} \left(\frac{-1}{d} \right) d^2, \\ S_8(n) &= 16 \cdot \begin{cases} \sum_{d|n} d^3 & \text{if } n \text{ is odd,} \\ \sum_{d|n} d^3 - 2 \sum_{d|g} d^3 & \text{if } n \text{ is even.} \end{cases} \end{aligned}$$

For example, $S_4(4) = 8(1 + 2) = 24$ and, indeed

$$\begin{aligned} 4 &= (\pm 1)^2 + (\pm 1)^2 + (\pm 1)^2 + (\pm 1)^2 = (\pm 2)^2 + 0 + 0 + 0 \\ &= 0 + (\pm 2)^2 + 0 + 0 = 0 + 0 + (\pm 2)^2 + 0 = 0 + 0 + 0 + (\pm 2)^2. \end{aligned}$$

So there are $16 + 2 + 2 + 2 + 2 = 24$ possible representations of the number 4 as a sum of 4 squares. Notice that $S_4(2) = S_4(4)$. In how many ways can 4 be represented as a sum of 6 squares? We write $4 = 2^2 \cdot 1$, so $\nu = 2$ and $g = 1$, and thus,

$$S_6(4) = \left(\left(\frac{-1}{1} \right) 2^{2\cdot 2+4} - 4 \right) \left(\left(\frac{-1}{1} \right) \cdot 1^2 \right) = (2^8 - 4) \cdot 1 = 252.$$

The formulas for $S_k(n)$ given above are derived using the theory of modular forms, as follows. We define a formal power series $\Theta(q)$ by

$$\Theta(q) = \sum_{j=-\infty}^{\infty} q^{j^2}$$

and, for $k \geq 2$, consider the power series expansion of the kth power of Θ:

$$\begin{aligned}(\Theta(q))^k &= \left(\sum_{j=-\infty}^{\infty} q^{j^2}\right)^k \\ &= \left(\sum_{a_1=-\infty}^{\infty} q^{a_1^2}\right)\cdots\left(\sum_{a_k=-\infty}^{\infty} q^{a_k^2}\right) = \sum_{n\geq 0} c_n q^n.\end{aligned}$$

What is the nth coefficient, c_n, of Θ^k? If the readers stare at the previous equation for a while, they will find that c_n is given by

$$c_n = \#\{(a_1, \ldots, a_k) \in \mathbb{Z}^k : a_1^2 + \cdots + a_k^2 = n\}.$$

Therefore, $c_n = S_k(n)$ and $(\Theta(q))^k = \sum_{n\geq 0} S_k(n)q^n$. In other words, Θ^k is a generating function for $S_k(n)$. But, how do we find closed formulas for $S_k(n)$? This is where the theory of modular forms becomes particularly useful, for it provides an alternative description of the coefficients of Θ^k.

It turns out that, for even $k \geq 2$, the function Θ^k is a modular form of weight $\frac{k}{2}$ (more precisely, it is a modular form for the group $\Gamma_1(4)$), and the space of all modular forms of weight $\frac{k}{2}$, denoted by $M_{\frac{k}{2}}(\Gamma_1(4))$, is finite dimensional (we will carefully define all these terms later). For instance, let $k = 4$. Then $M_2(\Gamma_1(4))$, the space of modular forms of weight $\frac{4}{2} = 2$ for $\Gamma_1(4)$, is a 2-dimensional $\mathbb{C}$-vector space and a basis is given by modular forms with q-expansions:

$$\begin{aligned}f(q) &= 1 + 24q^2 + 24q^4 + 96q^6 + 24q^8 + 144q^{10} + 96q^{12} + \cdots \\ g(q) &= q + 4q^3 + 6q^5 + 8q^7 + 13q^9 + 12q^{11} + 14q^{13} + \cdots .\end{aligned}$$

Therefore, $\Theta^4(q) = \lambda f(q) + \mu g(q)$ for some constants $\lambda, \mu \in \mathbb{C}$. We may compare q-expansions to find the values of λ and μ:

$$\begin{aligned} \Theta^4(q) = \sum_{n\geq 0} S_4(n) q^n &= 1 + 8q + 24q^2 + 32q^3 + 24q^4 + \cdots \\ \lambda f(q) + \mu g(q) &= \lambda + \mu q + 24\lambda q^2 + 4\mu q^3 + \cdots . \end{aligned}$$

Therefore, it is clear that $\lambda = 1$ and $\mu = 8$, so $\Theta^4 = f + 8g$. Since the expansions of f and g are easy to calculate (for example, using Sage; see Appendix A.2), we can easily calculate the coefficients of the q-expansion of Θ and, therefore, values of $S_4(n)$.

The exact formulas given above for $S_k(n)$, however, follow from some deeper facts. Here is a sketch of the ideas involved (the reader may skip these details for now and return here after reading Chapter 4): given $\Theta^4 = \sum c_n q^n$ and $F(q) = \sum(\sum_{d|n} d)q^n$, one can find an eigenvector $G(q) = \sum b_n q^n$ for a collection of linear maps T_n (the so-called Hecke operators, $T_n : M_2(\Gamma_1(4)) \to M_2(\Gamma_1(4))$) among spaces of modular forms, i.e., $T_n(G) = \lambda_n G$ for $n > 1$, and the eigenvalues $\lambda_n = b_n/b_1 = \sum_{d|n} d$. Moreover, the eigenvector G can be written explicitly as a combination of Θ^4 and F. Finally, one can show that the coefficients c_n must be given by the formula $c_n = 8\sum_{d|n,\ 4\nmid d} d$ (see [**Kob93**], III, §5, for more details). ■

1.3. L-functions

An L-function is a function $L(s)$, usually given as an infinite series of the form

$$L(s) = \sum_{n=1}^{\infty} a_n n^{-s} = \sum_{n=1}^{\infty} \frac{a_n}{n^s} = a_1 + \frac{a_2}{2^s} + \frac{a_3}{3^s} + \cdots$$

with some coefficients $a_n \in \mathbb{C}$. Typically, the function $L(s)$ converges for all complex numbers s in some half-plane (i.e., those s with real part larger than some constant), and in many cases $L(s)$ has an analytic or meromorphic continuation to the whole complex plane. Mathematicians are interested in L-functions because they are objects from analysis that, sometimes, capture very interesting algebraic information.

Example 1.3.1 (The Riemann zeta function)**.** The Riemann zeta function, usually denoted by $\zeta(s)$, is perhaps the most famous L-function:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots .$$

The reader may already know some values of ζ. For example $\zeta(2) = \sum \frac{1}{n^2}$ is convergent by the p-series test, and its value is $\pi^2/6$ (this value can be computed using Fourier analysis and Parseval's equality). The connection between $\zeta(s)$ and number theory comes from the fact that $\zeta(s)$ has an *Euler product*:

$$\begin{aligned} \zeta(s) &= \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}} \\ &= \left(\frac{1}{1-2^{-s}}\right) \cdot \left(\frac{1}{1-3^{-s}}\right) \cdot \left(\frac{1}{1-5^{-s}}\right) \cdots . \end{aligned}$$

This Euler product is not difficult to establish (Exercise 1.4.8) and has the very interesting consequence that any information on the distribution of the zeros of $\zeta(s)$ can be translated into information about the distribution of prime numbers among the natural numbers. ■

Example 1.3.2 (Dirichlet L-function)**.** *Let $a, N \in \mathbb{N}$ be relatively prime integers. Are there infinitely many primes p of the form $a+kN$ (i.e., $p \equiv a \mod N$) for $k \geq 0$?* The answer is *yes* and this fact, known as Dirichlet's theorem on primes in arithmetic progressions, was first proved by Dirichlet using a particular kind of L-function that we know today as a Dirichlet L-function.

Let $N > 0$. A *Dirichlet character* (modulo N) is a function $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ that is a homomorphism of groups, i.e., $\chi(nm) = \chi(n)\chi(m)$ for all $n, m \in (\mathbb{Z}/N\mathbb{Z})^\times$. Notice that $\chi(n) \in \mathbb{C}$ and $\chi(n)^{\varphi(N)} = 1$ for all $\gcd(n, N) = 1$. Therefore, $\chi(n)$ must be a root of unity. We extend χ to $\mathbb{Z}$ as follows. Let $a \in \mathbb{Z}$. If $\gcd(a, N) = 1$, then $\chi(a) = \chi(a \bmod N)$. Otherwise, if $\gcd(a, N) \neq 1$, then $\chi(a) = 0$.

A Dirichlet L-function is a function of the form

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s},$$

Figure 4. Johann Peter Gustav Lejeune Dirichlet (1805-1859) and Georg Friedrich Bernhard Riemann (1826-1866).

where χ is a given Dirichlet character. For example, one can take χ_0 to be the trivial Dirichlet character, i.e., $\chi_0(n) = 1$ for all $n \geq 1$. Then $L(s, \chi_0)$ is the Riemann zeta function $\zeta(s)$. Dirichlet L-functions also have Euler products:

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_p \frac{1}{1 - \chi(p)p^{-s}}.$$

The idea of the proof of Dirichlet's theorem generalizes the following proof, due to Euler, of the infinitude of the primes. Consider $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \frac{1}{1-p^{-s}}$ and suppose there are only finitely many primes. Then the product over all primes is finite, and therefore its value at $s = 1$ would be finite (a rational number, in fact). However, $\zeta(1) = \sum_{n=1}^{\infty} 1/n$ is the harmonic series, which diverges! Therefore, there must be infinitely many prime numbers.

Dirichlet adapted this argument by looking instead at a different function:

$$\Psi_{a,N}(s) = \sum_{p \equiv a \bmod N} \frac{1}{p^s}.$$

He showed that (a) for every non-trivial Dirichlet character χ modulo N, we have $L(1,\chi) \neq 0$ or ∞, and (b) this implies that $\Psi_{a,N}(1)$ diverges to ∞. Part (b) follows from the equality

$$\log(\zeta(s)) + \sum_{\substack{\chi \bmod N \\ \chi \neq 1}} \chi(a)^{-1} \log(L(s,\chi))$$
$$= \phi(N) \left(\sum_{p \equiv a \bmod N} \frac{1}{p^s} \right) + g(s),$$

where $g(s)$ is a function with $g(1)$ finite, and ϕ is the Euler ϕ-function. Therefore, there cannot be a finite number of primes of the form $p \equiv a \bmod N$. ∎

Example 1.3.3 (Representations of integers as sums of squares)**.** *Is the number $n > 0$ a sum of three integer squares?* In Subsection 1.2, we saw formulas for the number of representations of an integer as a sum of $k = 2$, 4, 6 and 8 integer squares, but we avoided the same question for odd k. The known formulas for $S_3(n)$, $S_5(n)$ and $S_7(n)$ involve values of Dirichlet L-functions.

Let us first define the Dirichlet character that we shall use here. The reader should be familiar with the Legendre symbol $\left(\frac{n}{p}\right)$, which is equal to 0 if $p|n$, equal to 1 if n is a square mod p, and equal to -1 if n is not a square mod p. Let $m > 0$ be a natural number with prime factorization $m = \prod_i p_i$ (the primes are not necessarily distinct). First we define

$$\left(\frac{n}{2}\right) = \begin{cases} 0 & \text{if } n \text{ is even,} \\ 1 & \text{if } n \equiv \pm 1 \bmod 8, \\ -1 & \text{if } n \equiv \pm 3 \bmod 8. \end{cases}$$

Now we are ready to define the *Kronecker symbol* of n over $m > 0$ by

$$\left(\frac{n}{m}\right) = \prod_i \left(\frac{n}{p_i}\right).$$

For any $n > 0$, the symbol $\left(\frac{-n}{\cdot}\right)$ induces a Dirichlet character χ_n defined by $\chi_n(a) = \left(\frac{-n}{a}\right)$, and we can define the associated L-function

by

$$L(s, \chi_n) = \sum_{a=1}^{\infty} \frac{\chi_n(a)}{a^s}.$$

We are ready to write down the formula for $S_3(n)$, due to Gauss, Dirichlet and Shimura (there are also formulas for $S_5(n)$, due to Eisenstein, Smith, Minkowski and Shimura, and a formula for $S_7(n)$, also due to Shimura). For simplicity, let us assume that n is odd and square free (for the utmost generality, please check **[Shi02]**):

$$S_3(n) = \begin{cases} 0 & \text{if } n \equiv 7 \bmod 8, \\ \frac{24\sqrt{n}}{\pi} L(1, \chi_n) & \text{otherwise.} \end{cases}$$

The reader is encouraged to investigate this problem further by attempting Exercises 1.4.6 and 1.4.7. ■

1.4. Exercises

Exercise 1.4.1. Use the divisibility properties of integers to show that the only solutions to $y^2 = x(x+1)(x+2)$ with $x, y \in \mathbb{Z}$ are $(0,0)$, $(-1,0)$ and $(-2,0)$. (Hint: If a and b are relatively prime and ab is a square, then a is a square and b is a square.)

Exercise 1.4.2. Find all the Pythagorean triples (a,b,c), i.e., $a, b, c \in \mathbb{Z}$ and $a^2 + b^2 = c^2$, such that $b^2 + c^2 = d^2$ for some $d \in \mathbb{Z}$. In other words, find all the integers a, b, c, d such that (a,b,c) and (b,c,d) are both Pythagorean triples. (Hint: You may assume that $y^2 = x(x+1)(x+2)$ has no rational points other than $(0,0)$, $(-1,0)$ and $(-2,0)$.)

Exercise 1.4.3. Prove Proposition 1.1.3; i.e., show that $f((a,b,c))$ is a point in E_n, that $g((x,y))$ is a triangle in C_n and that $f(g((x,y))) = (x,y)$ and $g(f((a,b,c))) = (a,b,c)$.

Exercise 1.4.4. Calculate $S_4(n)$, for $n = 1, 3, 5, 6$, by hand, using Jacobi's formula and also by finding all possible ways of writing n as a sum of 4 squares.

Exercise 1.4.5. The goal of this problem is to find the q-expansion of $\Theta^6(q)$:

(1) Find by hand the values of $S_6(n)$, for $n = 0, 1, 2$; i.e., find all possible ways to write $n = 0, 1, 2$ as a sum of 6 squares.

(2) Using Sage, calculate the dimension of $M_{\frac{k}{2}}(\Gamma_1(4))$ (see Appendix A.2) and a basis of modular forms for $k = 6$.

(3) Write Θ^6 as a linear combination of the basis elements found in part 2.

(4) Use part 3 to write the q-expansion of Θ^6 up to $O(q^{20})$.

(5) Use the expansion of Θ^6 to verify that $S_6(4) = 252$. Also, calculate $S_6(19)$ using Jacobi's formula and verify that it coincides with the coefficient of Θ^6 in front of the q^{19} term.

Exercise 1.4.6. Show that any integer $n \equiv 7 \bmod 8$ cannot be represented as a sum of three integer squares.

Exercise 1.4.7. Find the number of representations of $n = 3$ as a sum of 3 squares. Then compare your result with the value of the formula given in Example 1.3.3; i.e., use a computer to approximate

$$S_3(3) = \frac{24\sqrt{3}}{\pi} L(1, \chi_3) = \frac{24\sqrt{3}}{\pi} \sum_{a=1}^{\infty} \frac{\left(\frac{-3}{a}\right)}{a}$$

by adding the first $10,000$ terms of $L(1, \chi_3)$. Do the same for $n = 5$ and $n = 11$. Does the formula seem to work for $n = 2$? (*Note: the command* `kronecker(-n,m)` *calculates the Kronecker symbol* $\left(\frac{-n}{m}\right)$ *in* Sage.)

Exercise 1.4.8. Prove that the Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ has an Euler product; i.e., prove the following formal equality of series

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1 - p^{-s}}.$$

(Hint: There are two possible approaches:

Hint (a). Expand the right-hand side using the Fundamental Theorem of Arithmetic and the algebraic equality $\frac{1}{1+x} = \sum_{k=0}^{\infty} x^k$. [This approach helps build an intuition about what is going on, but may be hard to write into a rigorous proof]

Hint (b). Calculate $(1 - 1/2^s)\zeta(s)$ and $(1 - 1/3^s)(1 - 1/2^s)\zeta(s)$, etc.)

Chapter 2

Elliptic curves

In this chapter we summarize the main aspects of the theory of elliptic curves[1]. Unfortunately, we will not be able to provide many of the proofs because they are beyond the scope of this course. If the reader is not familiar with projective geometry or needs to refresh the memory, it is a good time to look at Appendix C or another reference (for example, [**SK52**] is a beautiful book on projective geometry).

2.1. Why elliptic curves?

A *Diophantine equation* is an equation given by a polynomial with integer coefficients, i.e.

$$f(x_1, x_2, \ldots, x_r) = 0 \tag{2.1}$$

with $f(x_1, \ldots, x_r) \in \mathbb{Z}[x_1, \ldots, x_r]$. Since antiquity, many mathematicians have studied the solutions in integers of Diophantine equations that arise from a variety of problems in number theory, e.g. $y^2 = x^3 - n^2 x$ is the Diophantine equation related to the study of the congruent number problem (see Example 1.1.2).

Since we would like to systematically study the integer solutions of Diophantine equations, we ask ourselves three basic questions:

[1]The contents of this chapter are largely based on the article [**Loz05**], in Spanish.

(a) Can we determine if Eq. (2.1) has any integral solutions, $x_i \in \mathbb{Z}$, or rational solutions, $x_i \in \mathbb{Q}$?

(b) If so, can we find any of the integral or rational solutions?

(c) Finally, can we find *all* solutions and prove that we have found all of them?

The first question was proposed by David Hilbert: *to devise a process according to which it can be determined in a finite number of operations whether the equation is solvable in rational integers.* This was Hilbert's tenth problem out of 23 fundamental questions that he proposed to the mathematical community during the Second International Congress of Mathematicians in Paris in the year 1900. Surprisingly, in 1970, Matiyasevich, Putnam and Robinson discovered that there is no such general algorithm that decides whether equation (2.1) has integer solutions (see [**Mat93**]). However, if we restrict our attention to certain particular cases, then we can answer questions (a), (b) and (c) posed above. The most significant advances have been obtained in equations with one and two variables:

- *Polynomials in one variable*:

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n = 0$$

with $a_i \in \mathbb{Z}$. This case is fairly simple. The following criterion determines how to search for rational or integral roots of a polynomial: if $\frac{p}{q} \in \mathbb{Q}$ is a solution of $f(x) = 0$, then a_n is divisible by p and a_0 is divisible by q.

- *Linear equations in two variables*:

$$ax + by = d$$

with $a, b, d \in \mathbb{Z}$ and $ab \neq 0$. Clearly, this type of equation always has an infinite number of rational solutions. As for integral solutions, Euclid's algorithm (to find $\gcd(a, b)$) determines if there are solutions $x, y \in \mathbb{Z}$ and, if so, produces all solutions. In particular, the equation has integral solutions if and only if d is divisible by $\gcd(a, b)$.

- *Quadratic equations (conics)*:

$ax^2 + bxy + cy^2 + dx + ey = f$ with $a, b, c, d, e, f \in \mathbb{Z}$.

Finding integral and rational points on a conic is a classical problem. Legendre's criterion determines whether there are rational solutions: a conic C has rational solutions if and only if C has points over $\mathbb{R}$ and over $\mathbb{Q}_p$, the p-adics, for all primes $p \geq 2$ (see Appendix D for a brief introduction to the p-adics). Essentially, Legendre's criterion says that the conic has rational solutions if and only if there are solutions modulo p^n for all primes p and all $n \geq 1$ but, in practice, one only needs to check this for a finite number of primes that depends on the coefficients of the conic.

If C has rational points, and we have found at least one point, then we can find all the rational solutions using a *stereographic projection* (see Exercise 2.12.2). The integral points on C, however, are much more difficult to find. The problem is equivalent to finding integral solutions to *Pell's equation* $x^2 - Dy^2 = 1$. There are several methods to solve Pell's equation. For example, one can use continued fractions (certain convergents $\frac{x}{y}$ of the continued fraction for $\sqrt{D}$ are integral solutions (x, y) of Pell's equation; see Exercise 2.12.2).

- *Cubic equations*:

$$aX^3 + bX^2Y + cXY^2 + dY^3 + eX^2 + fXY + gY^2 + hX + jY + k = 0.$$

A cubic equation in two variables may have no rational solutions, only 1 rational solution, a finite number of solutions, or infinitely many solutions. Unfortunately, we do not know any algorithm that yields all rational solutions of a cubic equation, although there are *conjectural* algorithms. In this chapter we will concentrate on this type of equation: a nonsingular cubic, i.e., no self-intersections or pinches, with at least one rational point (which will be our definition of an elliptic curve).

- *Higher degree.* Typically, curves defined by an equation of degree ≥ 4 have a genus ≥ 2 (but some equations of degree 4 have genus 1; see Example 2.2.5 and Exercise 2.12.4). The genus is an invariant that classifies curves according to their topology. Briefly, if we consider a curve as defined over $\mathbb{C}$,

then $C(\mathbb{C})$ may be considered as a surface over $\mathbb{R}$, and the genus of C counts the number of holes in the surface. For example, the projective line $\mathbb{P}^1(\mathbb{C})$ has no holes and $g = 0$ (the projective plane is homeomorphic to a sphere; see Appendix C for a quick introduction to projective geometry), and an elliptic curve has genus 1 (homeomorphic to a torus; see Theorem 3.2.5). Surprisingly, the genus of a curve is intimately related with the arithmetic of its points. More precisely, Louis Mordell conjectured that a curve C of genus ≥ 2 can only have a finite number of rational solutions. The conjecture was proved by Faltings in 1983.

2.2. Definition

Definition 2.2.1. An *elliptic curve* over $\mathbb{Q}$ is a smooth cubic projective curve E defined over $\mathbb{Q}$ with at least one rational point $\mathcal{O} \in E(\mathbb{Q})$ that we call the *origin*.

In other words, an elliptic curve is a curve E in the projective plane (see Appendix C) given by a cubic polynomial $F(X, Y, Z) = 0$ with rational coefficients, i.e.,

$$\begin{aligned} F(X,Y,Z) &= aX^3 + bX^2Y + cXY^2 + dY^3 \\ &\quad + eX^2Z + fXYZ + gY^2Z \\ &\quad + hXZ^2 + jYZ^2 + kZ^3 = 0, \end{aligned} \tag{2.2}$$

with coefficients $a, b, c, \ldots \in \mathbb{Q}$, and such that E is smooth; i.e., the tangent vector $\left(\frac{\partial F}{\partial X}(P), \frac{\partial F}{\partial Y}(P), \frac{\partial F}{\partial Z}(P)\right)$ does not vanish at any $P \in E$ (see Appendix C.5 for a brief introduction to singularities and nonsingular or smooth curves). If the coefficients $a, b, c, \ldots$ are in a field K, then we say that E is defined over K (and write E/K).

Even though the fact that E is a projective curve is crucial, we usually consider just affine charts of E, e.g. those points of the form $\{[X, Y, 1]\}$, and study instead the affine curve given by

$$\begin{aligned} &aX^3 + bX^2Y + cXY^2 + dY^3 \\ &+eX^2 + fXY + gY^2 + hX + jY + k = 0 \end{aligned} \tag{2.3}$$

but with the understanding that in this new model we may have left out some points of E *at infinity* (i.e., those points $[X, Y, 0]$ satisfying Eq. 2.2).

In general, one can find a change of coordinates that simplifies Eq. 2.3 enormously:

Proposition 2.2.2. *Let E be an elliptic curve, given by Eq. 2.2, defined over a field K of characteristic different from 2 or 3. Then there exists a curve $\widehat{E}$ given by*

$$zy^2 = x^3 + Axz^2 + Bz^3, \quad A, B \in K \quad \text{with} \quad 4A^3 + 27B^2 \neq 0$$

and an invertible change of variables $\psi : E \to \widehat{E}$ of the form

$$\psi([X, Y, Z]) = \left[\frac{f_1(X, Y, Z)}{g_1(X, Y, Z)}, \frac{f_2(X, Y, Z)}{g_2(X, Y, Z)}, \frac{f_3(X, Y, Z)}{g_3(X, Y, Z)}\right]$$

where f_i and g_i are polynomials with coefficients in K for $i = 1, 2, 3$, and the origin $\mathcal{O}$ is sent to the point $[0, 1, 0]$ of $\widehat{E}$, i.e., $\psi(\mathcal{O}) = [0, 1, 0]$.

The existence of such a change of variables is a consequence of the Riemann-Roch theorem of algebraic geometry (for a proof of the proposition see **[Sil86]**, Chapter III.3). The reference **[SiT92]**, Ch. I. 3, gives an explicit method to find the change of variables $\psi : E \to \widehat{E}$. See also pages 46-49 of **[Mil06]**.

A projective equation of the form $zy^2 = x^3 + Axz^2 + Bz^3$, or $y^2 = x^3 + Ax + B$ in affine coordinates, is called a *Weierstrass equation*. From now on, we will often work with an elliptic curve in this form. Notice that a curve E given by a Weierstrass equation $y^2 = x^3 + Ax + B$ is non-singular if and only if $4A^3 + 27B^2 \neq 0$, and it has a unique point at infinity, namely $[0, 1, 0]$, which we shall call the origin $\mathcal{O}$ or the point at infinity of E.

Sometimes we shall use a more general Weierstrass equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with $a_i \in \mathbb{Q}$ (we will explain the funky choice of notation for the coefficients later), but most of the time we will work with equations of the form $y^2 = x^3 + Ax + B$. It is easy to come up with a change of variables from one form to the other (see Exercise 2.12.3).

Example 2.2.3. Let $d \in \mathbb{Z}$, $d \neq 0$ and let E be the elliptic curve given by the cubic equation

$$X^3 + Y^3 = dZ^3$$

with $\mathcal{O} = [1, -1, 0]$. The reader should verify that E is a smooth curve. We wish to find a Weierstrass equation for E and, indeed, one can find a change of variables $\psi : E \to \widehat{E}$ given by

$$\psi([X, Y, Z]) = [12dZ,\ 36d(X - Y),\ X + Y] = [x, y, z]$$

such that $zy^2 = x^3 - 432d^2z^3$. The map ψ is invertible; the inverse map $\psi^{-1} : \widehat{E} \to E$ is

$$\psi^{-1}([x, y, z]) = \left[\frac{36dz + y}{72d},\ \frac{36dz - y}{72d},\ \frac{x}{12d}\right].$$

In affine coordinates, the change of variables is going from $X^3 + Y^3 = d$ to the curve $y^2 = x^3 - 432d^2$:

$$\begin{aligned} \psi(X, Y) &= \left(\frac{12d}{X + Y}, \frac{36d(X - Y)}{X + Y}\right), \\ \psi^{-1}(x, y) &= \left(\frac{36d + y}{6x}, \frac{36d - y}{6x}\right). \end{aligned}$$

■

Definition 2.2.4. Let $E : f(x, y) = 0$ be an elliptic curve with origin $\mathcal{O}$, and let $E' : g(X, Y) = 0$ be an elliptic curve with origin $\mathcal{O}'$. We say that E and E' are *isomorphic over* $\mathbb{Q}$ if there is an invertible change of variables $\psi : E \to E'$, defined by rational functions with coefficients in $\mathbb{Q}$, such that $\psi(\mathcal{O}) = \mathcal{O}'$.

Example 2.2.5. Sometimes, a curve given by a quartic polynomial can be isomorphic over $\mathbb{Q}$ to another curve given by a cubic polynomial. For instance, consider the curves

$$C/\mathbb{Q} : V^2 = U^4 + 1 \quad \text{and} \quad E/\mathbb{Q} : y^2 = x^3 - 4x.$$

The map $\psi : C \to E$ given by

$$\psi(U, V) = \left(\frac{2(V + 1)}{U^2}, \frac{4(V + 1)}{U^3}\right)$$

is an invertible rational map, defined over $\mathbb{Q}$, that sends $(0,1)$ to $\mathcal{O}$, and $\psi(0,-1) = (0,0)$. See Exercise 2.12.4. More generally, any quartic

$$C : V^2 = aU^4 + bU^3 + cU^2 + dU + q^2$$

for some $a, b, c, d, q \in \mathbb{Z}$ is isomorphic over $\mathbb{Q}$ to a curve of the form $E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$, also defined over $\mathbb{Q}$. The isomorphism is given in **[Was08]**, Theorem 2.17, p. 37.

Let E be an elliptic curve over $\mathbb{Q}$ given by a Weierstrass equation

$$E\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \quad a_i \in \mathbb{Q}.$$

With a change of variables $(x, y) \mapsto (u^{-2}x,\ u^{-3}y)$, we can find the equation of an elliptic curve isomorphic to E given by

$$y^2 + (a_1u)xy + (a_3u^3)y = x^3 + (a_2u^2)x^2 + (a_4u^4)x + (a_6u^6)$$

with coefficients $a_iu^i \in \mathbb{Z}$ for $i = 1, 2, 3, 4, 6$. By the way, *this* is one of the reasons for the peculiar numbering of the coefficients a_i.

Example 2.2.6. Let E be given by $y^2 = x^3 + \frac{x}{2} + \frac{5}{3}$. We may change variables by $x = \frac{X}{6^2}$ and $y = \frac{Y}{6^3}$ to obtain a new equation $Y^2 = X^3 + 648X + 77760$ with integral coefficients. ■

2.3. Integral points

In 1929, Siegel proved the following result about integral points $E(\mathbb{Z})$, i.e., about those points on E with integer coordinates:

Theorem 2.3.1 (Siegel's theorem; **[Sil86]**, Ch. IX, Thm. 3.1)**.** *Let $E/\mathbb{Q}$ be an elliptic curve given by $y^2 = x^3 + Ax + B$, with $A, B \in \mathbb{Z}$. Then E has only a finite number of integral points.*

Siegel's theorem is a consequence of a well-known theorem of Roth on Diophantine approximation. Unfortunately, Siegel's theorem is not effective and provides neither a method to find the integral points on E nor a bound on the number of integral points. However, in **[Bak90]**, Alan Baker found an alternative proof that provides an explicit upper bound on the size of the coefficients of an integral solution. More concretely, if $x, y \in \mathbb{Z}$ satisfy $y^2 = x^3 + Ax + B$, then

$$\max(|x|, |y|) < \exp((10^6 \cdot \max(|A|, |B|))^{10^6}).$$

Obviously, Baker's bound is not a very sharp bound, but it is theoretically interesting nonetheless.

2.4. The group structure on $E(\mathbb{Q})$

From now on, we will concentrate on trying to find all rational points on a curve $E : y^2 = x^3 + Ax + B$. We will use the following notation for the rational points on E:

$$E(\mathbb{Q}) = \{(x, y) \in E \mid x, y \in \mathbb{Q}\} \cup \{\mathcal{O}\}$$

where $\mathcal{O} = [0, 1, 0]$ is the point at infinity.

One of the aspects that makes the theory of elliptic curves so rich is that the set $E(\mathbb{Q})$ can be equipped with a group structure, geometric in nature. The (addition) operation on $E(\mathbb{Q})$ can be defined as follows (see Figure 1). Let E be given by a Weierstrass equation $y^2 = x^3 + Ax + B$ with $A, B \in \mathbb{Q}$. Let P and Q be two rational points in $E(\mathbb{Q})$ and let $\mathfrak{L} = \overline{PQ}$ be the line that goes through P and Q (if $P = Q$, then we define $\mathfrak{L}$ to be the tangent line to E at P). Since the curve E is defined by a cubic equation, and since we have defined $\mathfrak{L}$ so it already intersects E at two rational points, there must be a third point of intersection R in $\mathfrak{L} \cap E$, which is also defined over $\mathbb{Q}$, and

$$\mathfrak{L} \cap E(\mathbb{Q}) = \{P, Q, R\}.$$

The sum of P and Q, denoted by $P + Q$, is by definition the second point of intersection with E of the vertical line that goes through R, or in other words, the reflection of R across the x-axis.

It is easy to verify that the addition operation that we have defined on points of $E(\mathbb{Q})$ is commutative. The origin $\mathcal{O}$ is the zero element, and for every $P \in E(\mathbb{Q})$ there exists a point $-P$ such that $P + (-P) = \mathcal{O}$. If E is given by $y^2 = x^3 + Ax + B$ and $P = (x_0, y_0)$, then $-P = (x_0, -y_0)$. The addition is also associative (but this is not obvious, and it is tedious to prove) and, therefore, $(E, +)$ is an abelian group.

Example 2.4.1. Let E be the elliptic curve $y^2 = x^3 - 25x$, as in Example 1.1.2. The points $P = (5, 0)$ and $Q = (-4, 6)$ belong to

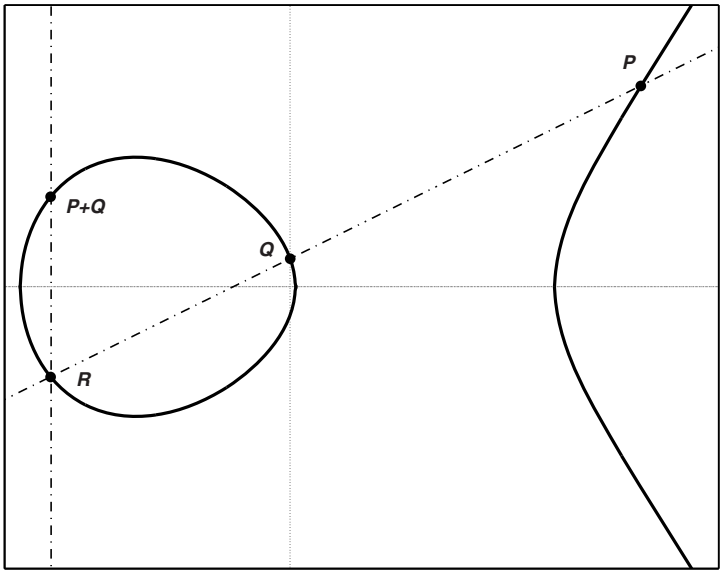


Figure 1. Addition of points on an elliptic curve

$E(\mathbb{Q})$. Let us find $P + Q$. First, we find the equation of the line $\mathfrak{L} = \overline{PQ}$. The slope must be

$$m = \frac{0-6}{5-(-4)} = -\frac{6}{9} = -\frac{2}{3}$$

and the line is $\mathfrak{L} : y = -\frac{2}{3}(x-5)$. Now we find the third point of intersection of $\mathfrak{L}$ and E by solving

$$\begin{cases} y = -\frac{2}{3}(x-5) \\ y^2 = x^3 - 25x. \end{cases}$$

Plugging the first equation into the second one, we obtain an equation

$$x^3 - \frac{4}{9}x^2 - \frac{185}{9}x - \frac{100}{9} = 0,$$

which factors as $(x-5)(x+4)(9x+5) = 0$. The first two factors are expected, since we already knew that $P = (5,0)$ and $Q = (-4,6)$ are in $\mathfrak{L} \cap E$. The third point of intersection must have $x = -\frac{5}{9}$, $y = -\frac{2}{3}(x-5) = \frac{100}{27}$ and, indeed, $R = (-\frac{5}{9}, \frac{100}{27})$ is a point in

$\mathfrak{L} \cap E(\mathbb{Q})$. Thus, $P + Q$ is the reflection of R across the x-axis, i.e., $P + Q = (-\frac{5}{9}, -\frac{100}{27})$.

Using Proposition 1.1.3, we may try to use the point $P + Q = (-\frac{5}{9}, -\frac{100}{27})$ to find a (new) right triangle with rational sides and area equal to 5, but this point corresponds to the triangle $(\frac{20}{3}, \frac{3}{2}, \frac{41}{6})$, the same triangle that corresponds to $Q = (-4, 6)$. In order to find a new triangle, let us find $Q + Q = 2Q$.

The line $\mathfrak{L}$ in this case is the tangent line to E at Q. The slope of $\mathfrak{L}$ can be found using implicit differentiation on $y^2 = x^3 - 25x$:

$$2y\frac{dy}{dx} = 3x^2 - 25, \quad \text{so} \quad \frac{dy}{dx} = \frac{3x^2 - 25}{2y}.$$

Hence, the slope of $\mathfrak{L}$ is $m = \frac{23}{12}$ and $\mathfrak{L} : y = \frac{23}{12}(x + 4) + 6$. In order to find R we need to solve

$$\begin{cases} y = \frac{23}{12}(x + 4) + 6 \\ y^2 = x^3 - 25x. \end{cases}$$

Simplifying yields $x^3 - \frac{529}{144}x^2 - \frac{1393}{18}x - \frac{1681}{9} = 0$, which factors as

$$(x + 4)^2(144x - 1681) = 0.$$

Once again, two factors were expected: $x = -4$ *needs* to be a double root because $\mathfrak{L}$ is *tangent* to E at $Q = (-4, 6)$. The third factor tells us that the x coordinate of R is $x = \frac{1681}{144}$, and $y = \frac{23}{12}(x+4)+6 = \frac{62279}{1728}$. Thus, $Q + Q = 2Q = (\frac{1681}{144}, -\frac{62279}{1728})$. This point corresponds to the right triangle

$$(a, b, c) = \left(\frac{1519}{492}, \frac{4920}{1519}, \frac{3344161}{747348}\right).$$

■

Example 2.4.2. Let $E : y^2 = x^3 + 1$ and put $P = (2, 3)$. Let us find P, $2P$, $3P$, etc.

- In order to find $2P$, first we need to find the tangent line to E at P, which is $y - 3 = 2(x - 2)$ or $y = 2x - 1$. The third point of intersection is $R = (0, -1)$, so $2P = (0, 1)$.
- To find $3P$, we add P and $2P$. The third point of intersection of E with the line that goes through P and $2P$ is $R' = (-1, 0)$; hence, $3P = (-1, 0)$.

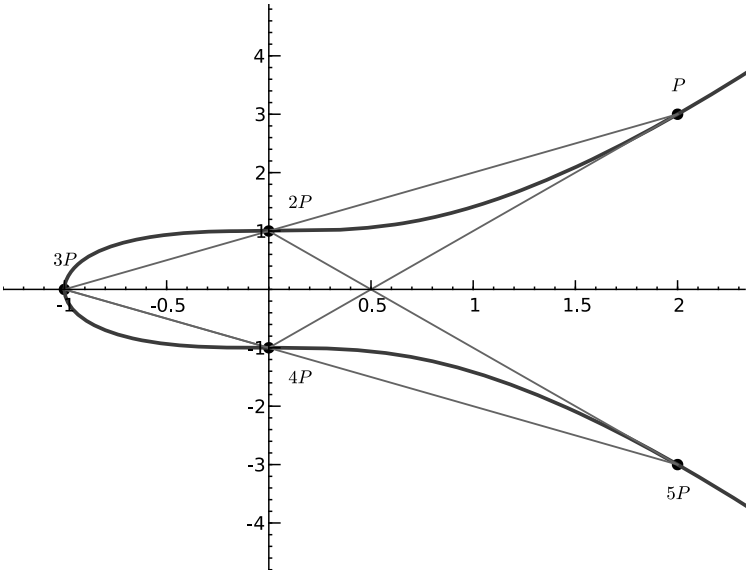


Figure 2. The rational points on $y^2 = x^3 + 1$, except the point at ∞.

- The point $4P$ can be found by adding $3P$ and P. The third point of intersection of E and the line through P and $3P$ is $R'' = 2P = (0, 1)$, and so $4P = P + 3P = (0, -1)$.
- We find $5P$ by adding $4P$ and P. Notice that the line that goes through $4P = (0, -1)$ and $P = (2, 3)$ is tangent at $(2, 3)$, so the third point of intersection is P. Thus, $5P = 4P + P = (2, -3)$.
- Finally, $6P = P + 5P$ but $5P = (2, -3) = -P$. Hence, $6P = P + (-P) = \mathcal{O}$, the point at infinity.

This means that P is a point of finite order, and its order equals 6. See Figure 2 (the Sage code for this graph can be found in the Appendix A.1.3). ■

The addition law can be defined more generally on any smooth projective cubic curve $E : f(X, Y, Z) = 0$, with a given rational point $\mathcal{O}$. Let $P, Q \in E(\mathbb{Q})$ and let $\mathfrak{L}$ be the line that goes through P and

Q. Let R be the third point of intersection of $\mathfrak{L}$ and E. Then R is also a rational point in $E(\mathbb{Q})$. Let $\mathfrak{L}'$ be the line through R and $\mathcal{O}$. We define $P+Q$ to be the third point of intersection of $\mathfrak{L}'$ and E. Notice that any vertical line $x=a$ in the affine plane passes through $[0,1,0]$, because the same line in projective coordinates is given by $x=az$ and $[0,1,0]$ belongs to such line. Thus, if E is given by a model $y^2=x^3+Ax+B$, and $\mathcal{O}$ is chosen to be the point $[0,1,0]$, then $\mathfrak{L}'$ is always a vertical line, so $P+Q$ is always the reflection of R with respect to the x axis.

The next step in the study of the structure of $E(\mathbb{Q})$ was conjectured by Jules Poincaré in 1908, proved by Louis Mordell in 1922 and generalized by André Weil in his thesis in 1928:

Theorem 2.4.3 (Mordell-Weil)**.** *$E(\mathbb{Q})$ is a finitely generated abelian group. In other words, there are points $P_1,\dots,P_n$ such that any other point Q in $E(\mathbb{Q})$ can be expressed as a linear combination*

$$Q=a_1P_1+a_2P_2+\cdots+a_nP_n$$

for some $a_i\in\mathbb{Z}$.

The group $E(\mathbb{Q})$ is usually called the Mordell-Weil group of E, in honor of the two mathematicians who proved the theorem.

Example 2.4.4. Consider the elliptic curve $E/\mathbb{Q}$ given by the Weierstrass equation

$$y^2+y=x^3-7x+6.$$

The set of rational points $E(\mathbb{Q})$ for this elliptic curve is infinite. For instance, the following points are on the curve:

$$(1,0),\ (2,0),\ (0,-3),\ (-3,-1),\ (8,-22),\ (-2,-4),\ (3,-4),$$
$$(3,3),\ (-1,-4),\ (1,-1),\ (0,2),\ (2,-1),\ (-2,3),\ (-1,3),$$
$$\left(\frac{1}{4},\frac{13}{8}\right),\ \left(\frac{25}{9},-\frac{91}{27}\right),\ \left(-\frac{26}{9},\frac{28}{27}\right),\ \left(\frac{7}{9},\frac{17}{27}\right),\dots.$$

At a first glance, it may seem very difficult to describe all the points on $E(\mathbb{Q})$, including those listed above, in a succinct manner. However, the Mordell-Weil theorem tells us that there must be a finite set of points that generate the whole group. Indeed, it can be proved that

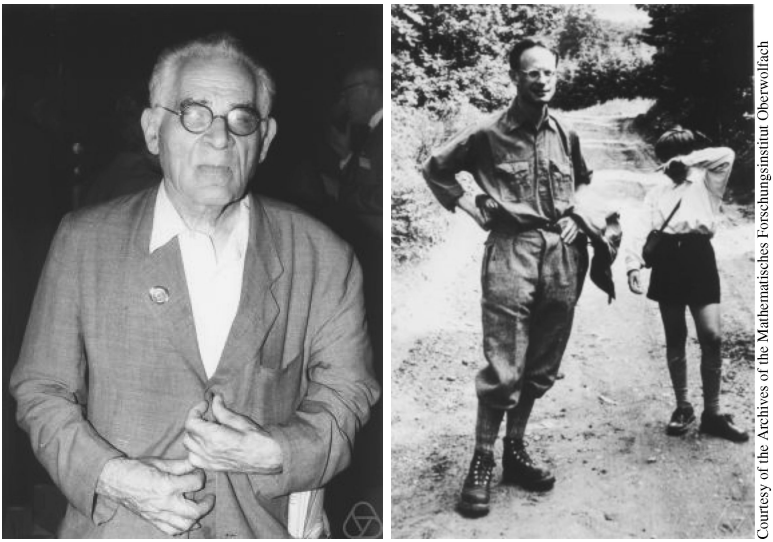


Figure 3. Louis Mordell (1888-1972) and André Weil (1906-1998).

the three points

$$P = (1, 0),\ Q = (2, 0),\quad \text{and}\quad R = (0, -3)$$

are generators of $E(\mathbb{Q})$. This means that *any other point* on $E(\mathbb{Q})$ can be expressed as a $\mathbb{Z}$-linear combination of P, Q and R. In other words,

$$E(\mathbb{Q}) = \{a \cdot P + b \cdot Q + c \cdot R : a, b, c \in \mathbb{Z}\}.$$

For instance,

$$(-3, -1) = P + Q,\ (8, -22) = P + R,\ (-2, -4) = P - Q,$$

$$(-1, -4) = Q - R \ \text{ and } (3, 3) = P - R.$$

The proof of the theorem has three fundamental ingredients: the so-called *weak* Mordell-Weil theorem $E(\mathbb{Q})/mE(\mathbb{Q})$ is finite for any $m \geq 2$; see below); the concept of height functions on abelian groups and the *descent theorem*, which establishes that an abelian group A with a height function h, such that A/mA is finite (for some $m \geq 2$), is finitely generated.

Theorem 2.4.5 (weak Mordell-Weil)**.** *$E(\mathbb{Q})/mE(\mathbb{Q})$ is a finite group for all $m \geq 2$.*

We will discuss the proof of a special case of the weak Mordell-Weil theorem in Section 2.9 (see Corollary 2.9.7).

It follows from the Mordell-Weil theorem and the general structure theory of finitely generated abelian groups that

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{torsion}} \oplus \mathbb{Z}^{R_E}. \tag{2.4}$$

In other words, $E(\mathbb{Q})$ is isomorphic to the direct sum of two abelian groups (notice however that this decomposition *is not* canonical!). The first summand is a finite group formed by all *torsion* elements, i.e., those points on E of finite order:

$$E(\mathbb{Q})_{\text{torsion}} = \{P \in E(\mathbb{Q}) : \text{ there is } n \in \mathbb{N} \text{ such that } nP = \mathcal{O}\}.$$

The second summand of Eq. (2.4), sometimes called the *free part*, is $\mathbb{Z}^{R_E}$, i.e., R_E copies of $\mathbb{Z}$ for some integer $R_E \geq 0$. It is generated by R_E points of $E(\mathbb{Q})$ of infinite order (i.e., $P \in E(\mathbb{Q})$ such that $nP \neq \mathcal{O}$ for all non-zero $n \in \mathbb{Z}$). The number R_E is called the *rank* of the elliptic curve $E/\mathbb{Q}$. Notice, however, that the set

$$F = \{P \in E(\mathbb{Q}) : P \text{ is of infinite order}\} \cup \{\mathcal{O}\}$$

is not a subgroup of $E(\mathbb{Q})$ if the torsion subgroup is non-trivial. For instance, if T is a torsion point and P is of infinite order, then P and $P+T$ belong to F but $T = (P+T) - P$ does not belong to F. This fact makes the isomorphism of Eq. (2.4) not canonical because the subgroup of $E(\mathbb{Q})$ isomorphic to $\mathbb{Z}^{R_E}$ cannot be chosen, in general, in a unique way.

Example 2.4.6. The following are some examples of elliptic curves and their Mordell-Weil groups:

(1) The curve $E_1/\mathbb{Q} : y^2 = x^3 + 6$ has no rational points, other than the point at infinity $\mathcal{O}$. Therefore, there are no torsion points (other than $\mathcal{O}$) and no points of infinite order. In particular, the rank is 0, and $E_1(\mathbb{Q}) = \{\mathcal{O}\}$.

(2) The curve $E_2/\mathbb{Q} : y^2 = x^3 + 1$ has only 6 rational points. As we saw in Example 2.4.2, the point $P = (2, 3)$ has exact order 6. Therefore $E_2(\mathbb{Q}) \cong \mathbb{Z}/6\mathbb{Z}$ is an isomorphism of

groups. Since there are no points of infinite order, the rank of $E_2/\mathbb{Q}$ is 0, and

$$E_2(\mathbb{Q}) = \{\mathcal{O},\ P,\ 2P,\ 3P,\ 4P,\ 5P\} = \{\mathcal{O}, (2, \pm 3), (0, \pm 1), (-1, 0)\}.$$

(3) The curve $E_3/\mathbb{Q} : y^2 = x^3 - 2$ does not have any rational torsion points other than $\mathcal{O}$ (as we shall see in the next section). However, the point $P = (3, 5)$ is a rational point. Thus, P must be a point of infinite order and $E_3(\mathbb{Q})$ contains infinitely many distinct rational points. In fact, the rank of E_3 is equal to 1 and P is a generator of all of $E_3(\mathbb{Q})$, i.e.,

$$E_3(\mathbb{Q}) = \{nP : n \in \mathbb{Z}\} \quad \text{and} \quad E_3(\mathbb{Q}) \cong \mathbb{Z}.$$

(4) The elliptic curve $E_4/\mathbb{Q} : y^2 = x^3 + 7105x^2 + 1327104x$ features both torsion and infinite order points. In fact, $E_4(\mathbb{Q}) \cong \mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}^3$. The torsion subgroup is generated by the point $T = (1152, 111744)$ of order 4. The free part is generated by three points of infinite order:

$$P_1 = (-6912, 6912),\ P_2 = (-5832, 188568),\ P_3 = (-5400, 206280).$$

Hence

$$E_4(\mathbb{Q}) = \{aT + bP_1 + cP_2 + dP_3 : a = 0, 1, 2 \text{ or } 3 \text{ and } b, c, d \in \mathbb{Z}\}.$$

As we mentioned above, the isomorphism $E_4(\mathbb{Q}) \cong \mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}^3$ is not canonical. For instance, $E_4(\mathbb{Q}) \cong \langle T \rangle \oplus \langle P_1, P_2, P_3 \rangle$ but also $E_4(\mathbb{Q}) \cong \langle T \rangle \oplus \langle P_1', P_2, P_3 \rangle$ with $P_1' = P_1 + T$. ■

The rank of $E/\mathbb{Q}$ is, in a sense, a measurement of the arithmetic complexity of the elliptic curve. It is not known if there is an upper bound for the possible values of R_E (the largest rank known is 28, discovered by Noam Elkies; see Andrej Dujella's website **[Duj09]** for up-to-date records and examples of curves with "high" ranks). It has been conjectured (with some controversy) that ranks can be arbitrarily large; i.e., for all $n \in \mathbb{N}$ there exists an elliptic curve E over $\mathbb{Q}$ with $R_E \geq n$. We state this as a conjecture for future reference:

Conjecture 2.4.7 (Conjecture of the rank)**.** *Let $N \geq 0$ be a natural number. Then there exists an elliptic curve E defined over $\mathbb{Q}$ with rank $R_E \geq N$.*

One of the key pieces of evidence in favor of such a conjecture was offered by Shafarevich and Tate, who proved that there exist elliptic curves defined over function fields $\mathbb{F}_p(T)$ and with arbitrarily large ranks ($\mathbb{F}_p(T)$ is a field that shares many similar properties with $\mathbb{Q}$; see **[ShT67]**). In any case, the problem of finding elliptic curves of high rank is particularly interesting because of its arithmetic and computational complexity.

2.5. The torsion subgroup

In this section we concentrate on the torsion points of an elliptic curve:

$$E(\mathbb{Q})_{\text{torsion}} = \{P \in E(\mathbb{Q}) : \text{ there is } n \in \mathbb{N} \text{ such that } nP = \mathcal{O}\}.$$

Example 2.5.1. The curve $E_n : y^2 = x^3 - n^2x = x(x-n)(x+n)$ has three obvious rational points, namely $P = (0,0), Q = (-n,0), T = (n,0)$, and it is easy to check (see Exercise 2.12.6) that each one of these points is torsion of order 2, i.e., $2P = 2Q = 2T = \mathcal{O}$, and $P + Q = T$. In fact $E_n(\mathbb{Q})_{\text{torsion}} = \{\mathcal{O}, P, Q, T\} \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$. ■

Note that the Mordell-Weil theorem implies that $E(\mathbb{Q})_{\text{torsion}}$ is always finite. This fact prompts a natural question: *what abelian groups can appear in this context?* The answer was conjectured by Ogg and proven by Mazur:

Theorem 2.5.2 (Ogg's conjecture; Mazur, **[Maz77]**, **[Maz78]**)**.** *Let $E/\mathbb{Q}$ be an elliptic curve. Then $E(\mathbb{Q})_{torsion}$ is isomorphic to one of the following groups:*

$$\begin{aligned} &\mathbb{Z}/N\mathbb{Z} \quad \textit{with} \quad 1 \le N \le 10 \textit{ or } N = 12, \textit{ or} \\ &\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2M\mathbb{Z} \quad \textit{with} \quad 1 \le M \le 4. \end{aligned} \tag{2.5}$$

Example 2.5.3. For instance, the torsion subgroup of the elliptic curve with Weierstrass equation $y^2 + 43xy - 210y = x^3 - 210x^2$ is isomorphic to $\mathbb{Z}/12\mathbb{Z}$ and it is generated by the point $(0, 210)$. The elliptic curve $y^2 + 17xy - 120y = x^3 - 60x^2$ has a torsion subgroup isomorphic to $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/8\mathbb{Z}$, generated by the rational points $(30, -90)$ and $(-40, 400)$. See Figure 4 for a complete list of examples with each possible torsion subgroup. ■

Curve	Torsion	Generators
$y^2 = x^3 - 2$	trivial	$\mathcal{O}$
$y^2 = x^3 + 8$	$\mathbb{Z}/2\mathbb{Z}$	$(-2, 0)$
$y^2 = x^3 + 4$	$\mathbb{Z}/3\mathbb{Z}$	$(0, 2)$
$y^2 = x^3 + 4x$	$\mathbb{Z}/4\mathbb{Z}$	$(2, 4)$
$y^2 - y = x^3 - x^2$	$\mathbb{Z}/5\mathbb{Z}$	$(0, 1)$
$y^2 = x^3 + 1$	$\mathbb{Z}/6\mathbb{Z}$	$(2, 3)$
$y^2 = x^3 - 43x + 166$	$\mathbb{Z}/7\mathbb{Z}$	$(3, 8)$
$y^2 + 7xy = x^3 + 16x$	$\mathbb{Z}/8\mathbb{Z}$	$(-2, 10)$
$y^2 + xy + y = x^3 - x^2 - 14x + 29$	$\mathbb{Z}/9\mathbb{Z}$	$(3, 1)$
$y^2 + xy = x^3 - 45x + 81$	$\mathbb{Z}/10\mathbb{Z}$	$(0, 9)$
$y^2 + 43xy - 210y = x^3 - 210x^2$	$\mathbb{Z}/12\mathbb{Z}$	$(0, 210)$
$y^2 = x^3 - 4x$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$	$\binom{(2,0)}{(0,0)}$
$y^2 = x^3 + 2x^2 - 3x$	$\mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$	$\binom{(3,6)}{(0,0)}$
$y^2 + 5xy - 6y = x^3 - 3x^2$	$\mathbb{Z}/6\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$	$\binom{(-3,18)}{(2,-2)}$
$y^2 + 17xy - 120y = x^3 - 60x^2$	$\mathbb{Z}/8\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$	$\binom{(30,-90)}{(-40,400)}$

Figure 4. Examples of each of the possible torsion subgroups over $\mathbb{Q}$.

Furthermore, it is known that, if G is any of the groups in Eq. 2.5, there are infinitely many elliptic curves whose torsion subgroup is isomorphic to G. See, for example, [**Kub76**], Table 3, p. 217. For the convenience of the reader, the table in Kubert's article is reproduced in Appendix E.

Example 2.5.4. Let $E_t : y^2 + (1-t)xy - ty = x^3 - tx^2$ with $t \in \mathbb{Q}$ and $\Delta_t = t^5(t^2 - 11t - 1) \neq 0$. Then, the torsion subgroup of $E_t(\mathbb{Q})$ contains a subgroup isomorphic to $\mathbb{Z}/5\mathbb{Z}$, and $(0, 0)$ is a point of exact order 5. Conversely, if $E : y^2 = x^3 + Ax + B$ is an elliptic curve with torsion subgroup equal to $\mathbb{Z}/5\mathbb{Z}$, then there is an invertible change of variables that takes E to an equation of the form E_t for some $t \in \mathbb{Q}$. See also Examples E.1.1 and E.1.2. ■

A useful and simple consequence of Mazur's theorem is that if the order of a rational point $P \in E(\mathbb{Q})$ is larger than 12, then P must be a point of infinite order and, therefore, $E(\mathbb{Q})$ contains an infinite

number of distinct rational points. Except for this criterion, Mazur's theorem is not very helpful in effectively computing the torsion subgroup of a given elliptic curve. However, the following result, proven independently by E. Lutz and T. Nagell, provides a simple algorithm to determine $E(\mathbb{Q})_{\text{torsion}}$:

Theorem 2.5.5 (Nagell-Lutz, [**Nag35**], [**Lut37**]). *Let $E/\mathbb{Q}$ be an elliptic curve with Weierstrass equation*

$$y^2 = x^3 + Ax + B, \quad A, B \in \mathbb{Z}.$$

Then, every torsion point $P \neq \mathcal{O}$ of E satisfies:

(1) *The coordinates of P are integers, i.e., $x(P), y(P) \in \mathbb{Z}$.*

(2) *If P is a point of order $n \geq 3$, then $4A^3 + 27B^2$ is divisible by $y(P)^2$.*

(3) *If P is of order 2, then $y(P) = 0$ and $x(P)^3 + Ax(P) + B = 0$.*

For a proof, see [**Sil86**], Ch. VIII, Corollary 7.2, or [**Mil06**], Ch. II, Theorem 5.1.

Example 2.5.6. Let $E/\mathbb{Q} : y^2 = x^3 - 2$, so that $A = 0$ and $B = -2$. The polynomial $x^3 - 2$ does not have any rational roots, so $E(\mathbb{Q})$ does not contain any points of order 2. Also, $4A^3 + 27B^2 = 27 \cdot 4$. Thus, if $(x(P), y(P))$ are the coordinates of a torsion point in $E(\mathbb{Q})$, then $y(P)$ is an integer and $y(P)^2$ divides $27 \cdot 4$. This implies that $y(P) = \pm 1$, ± 2, ± 3, or ± 6. In turn, this implies that $x(P)^3 = 3$, 6, 11 or 38, respectively. However, $x(P)$ is an integer, and none of 3, 6, 11 or 38 are a perfect cube. Thus, $E(\mathbb{Q})_{\text{torsion}}$ is trivial (i.e., the only torsion point is $\mathcal{O}$).

Example 2.5.7. Let $p \geq 2$ be a prime number and let us define a curve $E_p : y^2 = x^3 + p^2$. Since $x^3 + p^2 = 0$ does not have any rational roots, $E_p(\mathbb{Q})$ does not contain points of order 2. Let P be a torsion point on $E_p(\mathbb{Q})$. The list of all squares dividing $4A^3 + 27B^2 = 27p^4$ is short, and by the Nagell-Lutz theorem the possible values for $y(P)$ are:

$$y = \pm 1, \ \pm p, \ \pm p^2, \ \pm 3p, \ \pm 3p^2, \text{ and } \pm 3.$$

Clearly, $(0, \pm p) \in E_p(\mathbb{Q})$ and one can show that those two points and $\mathcal{O}$ are the only torsion points; see Exercise 2.12.8. Thus, the torsion subgroup of $E_p(\mathbb{Q})$ is isomorphic to $\mathbb{Z}/3\mathbb{Z}$ for any prime $p \geq 2$. ■

2.6. Elliptic curves over finite fields

Let $p \geq 2$ be a prime and let $\mathbb{F}_p$ be the finite field with p elements, i.e.,

$$\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z} = \{a \bmod p : a = 0, 1, 2, \ldots, p-1\}.$$

$\mathbb{F}_p$ is a field and we may consider elliptic curves defined over $\mathbb{F}_p$. As for elliptic curves over $\mathbb{Q}$, there are two conditions that need to be satisfied: the curve needs to be given by a cubic equation, and the curve needs to be smooth.

Example 2.6.1. For instance, $E : y^2 \equiv x^3 + 1 \bmod 5$ is an elliptic curve defined over $\mathbb{F}_5$. It is clearly given by a cubic equation ($zy^2 \equiv x^3 + z^3 \bmod 5$ in the projective plane $\mathbb{P}^2(\mathbb{F}_5)$) and it is smooth, because for $F \equiv zy^2 - x^3 - z^3 \bmod 5$, the partial derivatives are:

$$\frac{\partial F}{\partial x} \equiv -3x^2, \quad \frac{\partial F}{\partial y} \equiv 2yz, \quad \frac{\partial F}{\partial z} \equiv y^2 - 3z^2 \bmod 5.$$

Thus, if the partial derivatives are congruent to 0 modulo 5, then $x \equiv 0 \bmod 5$ and $yz \equiv 0 \bmod 5$. The latter congruence implies that y or $z \equiv 0 \bmod 5$, and $\partial F/\partial z \equiv 0$ implies that $y \equiv z \equiv 0 \bmod 5$. Since $[0, 0, 0]$ is not a point in the projective plane, we conclude that there are no singular points on $E/\mathbb{F}_5$.

However, $C/\mathbb{F}_3 : y^2 \equiv x^3 + 1 \bmod 3$ is not an elliptic curve because it is not smooth. Indeed, the point $P = (2 \bmod 3, 0 \bmod 3) \in C(\mathbb{F}_3)$ is a singular point:

$$\begin{aligned} \frac{\partial F}{\partial x}(P) &\equiv -3 \cdot 2^2 \equiv 0, \quad \frac{\partial F}{\partial y}(P) \equiv 2 \cdot 0 \cdot 1 \equiv 0, \quad \text{and} \\ \frac{\partial F}{\partial z}(P) &\equiv 0^2 - 3 \cdot 1^2 \equiv 0 \bmod 3. \quad \blacksquare \end{aligned}$$

Let $E/\mathbb{Q}$ be an elliptic curve given by a Weierstrass equation $y^2 = x^3 + Ax + B$ with integer coefficients $A, B \in \mathbb{Z}$, and let $p \geq 2$ be a prime number. If we reduce A and B modulo p, then we obtain the equation of a curve $\widetilde{E}$ given by a cubic curve and defined over the field $\mathbb{F}_p$. Even though E is smooth as a curve over $\mathbb{Q}$, the curve $\widetilde{E}$ may be singular over $\mathbb{F}_p$. In the previous example, we saw that $E/\mathbb{Q} : y^2 = x^3 + 1$ is smooth over $\mathbb{Q}$ and $\mathbb{F}_5$ but it has a singularity over $\mathbb{F}_3$. If the reduction curve $\widetilde{E}$ is smooth, then it is an elliptic curve over $\mathbb{F}_p$.

Example 2.6.2. Sometimes the reduction of a model for an elliptic curve E modulo a prime p is not smooth, but it is smooth for some other models of E. For instance, consider the curve $E : y^2 = x^3 + 15625$. Then $\widetilde{E} \equiv E \bmod 5$ is not smooth over $\mathbb{F}_5$ because the point $(0, 0) \bmod 5$ is a singular point. However, using the invertible change of variables $(x, y) \mapsto (5^2X, 5^3Y)$, we obtain a new model over $\mathbb{Q}$ for E given by $E' : Y^2 = X^3 + 1$, which is smooth when we reduce it modulo 5. The problem here is that the model we chose for E is not *minimal*. We describe what we mean by minimal next. ■

Definition 2.6.3. Let E be an elliptic curve given by $y^2 = x^3 + Ax + B$, with $A, B \in \mathbb{Q}$.

(1) We define Δ_E, the *discriminant* of E, by

$$\Delta_E = -16(4A^3 + 27B^2).$$

For a definition of the discriminant for more general Weierstrass equations, see for example **[Sil86]**, p. 46.

(2) Let S be the set of all elliptic curves E' that are isomorphic to E over $\mathbb{Q}$ (see Definition 2.2.4) and such that the discriminant of E' is an integer. The *minimal discriminant* of E is the integer $\Delta_{E'}$ that attains the minimum of the set $\{|\Delta_{E'}| : E' \in S\}$. In other words, the minimal discriminant is the smallest integral discriminant (in absolute value) of an elliptic curve that is isomorphic to E over $\mathbb{Q}$. If E' is the model for E with minimal discriminant, we say that E' is a *minimal model* for E.

Example 2.6.4. The curve $E : y^2 = x^3 + 5^6$ has discriminant $\Delta_E = -2^4 3^3 5^{12}$, and the curve $E' : y^2 = x^3 + 1$ has discriminant $\Delta_{E'} = -2^4 3^3$. Since E and E' are isomorphic (see Definition 2.2.4 and Example 2.6.2), then Δ_E cannot be the minimal discriminant for E and $y^2 = x^3 + 5^6$ is not a minimal model. In fact, the minimal discriminant is $\Delta_{E'} = -432$ and E' is a minimal model. ■

Before we go on to describe the types of reduction modulo p that one can encounter, we need a little bit of background on types of singularities. Let $\widetilde{E}$ be a cubic curve over a field K with Weierstrass

equation $f(x,y)=0$, where

$$f(x,y)=y^2+a_1xy+a_3y-x^3-a_2x^2-a_4x-a_6,$$

and suppose that $\widetilde{E}$ has a singular point $P=(x_0,y_0)$, i.e., $\partial f/\partial x(P)=\partial f/\partial y(P)=0$. Thus, we can write the Taylor expansion of $f(x,y)$ around (x_0,y_0) as follows:

$$\begin{aligned} & f(x,y)-f(x_0,y_0) \\ = \;& \lambda_1(x-x_0)^2+\lambda_2(x-x_0)(y-y_0)+\lambda_3(y-y_0)^2-(x-x_0)^3 \\ = \;& ((y-y_0)-\alpha(x-x_0))\cdot((y-y_0)-\beta(x-x_0))-(x-x_0)^3 \end{aligned}$$

for some $\lambda_i\in K$ and $\alpha,\beta\in\overline{K}$ (an algebraic closure of K).

Definition 2.6.5. The singular point $P\in\widetilde{E}$ is a *node* if $\alpha\neq\beta$. In this case there are two different tangent lines to $\widetilde{E}$ at P, namely

$$y-y_0=\alpha(x-x_0),\quad y-y_0=\beta(x-x_0).$$

If $\alpha=\beta$, then we say that P is a *cusp*, and there is a unique tangent line at P.

Definition 2.6.6. Let $E/\mathbb{Q}$ be an elliptic curve given by a minimal model, let $p\geq 2$ be a prime and let $\widetilde{E}$ be the reduction curve of E modulo p. We say that $E/\mathbb{Q}$ has *good reduction* modulo p if $\widetilde{E}$ is smooth and hence is an elliptic curve over $\mathbb{F}_p$. If $\widetilde{E}$ is singular at a point $P\in E(\mathbb{F}_p)$, then we say that $E/\mathbb{Q}$ has bad reduction at p and we distinguish two cases:

(1) If $\widetilde{E}$ has a cusp at P, then we say that E has *additive (or unstable) reduction*.

(2) If $\widetilde{E}$ has a node at P, then we say that E has *multiplicative (or semistable) reduction*. If the slopes of the tangent lines (α and β as above) are in $\mathbb{F}_p$, then the reduction is said to be *split* multiplicative (and *non-split* otherwise).

Example 2.6.7. Let us see some examples of elliptic curves with different types of reduction:

(1) $E_1\colon y^2=x^3+35x+5$ has good reduction at $p=7$, because $y^2\equiv x^3+5 \bmod 7$ is a non-singular curve over $\mathbb{F}_7$.

(2) However E_1 has bad reduction at $p = 5$, and the reduction is additive, since modulo 5 we can write the equation as $((y-0) - 0 \cdot (x-0))^2 - x^3$ and the unique slope is 0.

(3) The elliptic curve $E_2 \colon y^2 = x^3 - x^2 + 35$ has bad multiplicative reduction at 5 and 7. The reduction at 5 is split, while the reduction at 7 is non-split. Indeed, modulo 5 we can write the equation as

$$((y-0) - 2(x-0)) \cdot ((y-0) + 2(x-0)) - x^3,$$

the slopes being 2 and -2. However, for $p = 7$, the slopes are not in $\mathbb{F}_7$ (because -1 is not a quadratic residue in $\mathbb{F}_7$). Indeed, when we reduce the equation modulo 7, we obtain

$$y^2 + x^2 - x^3 \bmod 7$$

and $y^2 + x^2$ can only be factored in $\mathbb{F}_7[i]$ but not in $\mathbb{F}_7$.

(4) Let E_3 be an elliptic curve given by the model $y^2 + y = x^3 - x^2 - 10x - 20$. This is a minimal model for E_3 and its (minimal) discriminant is $\Delta_{E_3} = -11^5$. The prime 11 is the unique prime of bad reduction and the reduction is split multiplicative. Indeed, the point $(5, 5)$ mod 11 is a singular point on $E_3(\mathbb{F}_{11})$ and

$$\begin{aligned} f(x,y) &= y^2 + y + x^2 + 10x + 20 - x^3 \\ &= (y - 5 - 5(x-5)) \cdot (y - 5 + 5(x-5)) - (x-5)^3. \end{aligned}$$

Hence, the slopes at $(5, 5)$ are 5 and -5, which are obviously in $\mathbb{F}_{11}$ and distinct.

■

Proposition 2.6.8. *Let K be a field and let E/K be a cubic curve given by $y^2 = f(x)$, where $f(x)$ is a monic cubic polynomial in $K[x]$. Suppose that $f(x) = (x-\alpha)(x-\beta)(x-\gamma)$ with $\alpha, \beta, \gamma \in \overline{K}$ (an algebraic closure of K) and put*

$$D = (\alpha - \beta)^2 (\alpha - \gamma)^2 (\beta - \gamma)^2.$$

Then E is non-singular if and only if $D \neq 0$.

The proof of the proposition is left as an exercise (see Exercise 2.12.9). Notice that the quantity D that appears in the previous

proposition is the *discriminant* of the polynomial $f(x)$. The discriminant of $E/\mathbb{Q}$, Δ_E as in Definition 2.6.3, is a multiple of D; in fact, $\Delta_E = 16D$. This fact together with Proposition 2.6.8 yield the following corollary:

Corollary 2.6.9. *Let $E/\mathbb{Q}$ be an elliptic curve with coefficients in $\mathbb{Z}$. Let $p \geq 2$ be a prime. If E has bad reduction at p, then $p \mid \Delta_E$. In fact, if E is given by a minimal model, then $p \mid \Delta_E$ if and only if E has bad reduction at p.*

Example 2.6.10. The discriminant of the elliptic curve $E_1 \colon y^2 = x^3 + 35x + 5$ of Example 2.6.7 is $\Delta_{E_1} = -2754800 = -2^4 \cdot 5^2 \cdot 71 \cdot 97$ (and, in fact, this is the minimal discriminant of E_1). Thus, E_1 has good reduction at 7 but it has bad reduction at 2, 5, 71 and 97. The reduction at 71 and 97 is multiplicative. ■

Let $\widetilde{E}$ be an elliptic curve defined over a finite field $\mathbb{F}_q$ with q elements, where $q = p^r$ and $p \geq 2$ is prime. Notice that $\widetilde{E}(\mathbb{F}_q) \subseteq \mathbb{P}^2(\mathbb{F}_q)$, and the projective plane over $\mathbb{F}_q$ only has a finite number of points (how many?). Thus, the number $N_q := |\widetilde{E}(\mathbb{F}_q)|$, i.e., the number of points on $\widetilde{E}$ over $\mathbb{F}_q$, is finite. The following theorem provides a bound for N_q. This result was conjectured by Emil Artin (in his thesis) and was proved by Helmut Hasse in the 1930's:

Theorem 2.6.11 (Hasse; **[Sil86]**, Ch. V, Theorem 1.1)**.** *Let $\widetilde{E}$ be an elliptic curve defined over $\mathbb{F}_q$. Then*

$$q + 1 - 2\sqrt{q} < N_q < q + 1 + 2\sqrt{q},$$

where $N_q = |\widetilde{E}(\mathbb{F}_q)|$.

Remark 2.6.12. Heuristically, we expect that N_q is approximately $q+1$, in agreement with Hasse's bound. Indeed, let $E/\mathbb{Q}$ be an elliptic curve given by $y^2 = x^3 + Ax + B$, with $A, B \in \mathbb{Z}$, and let $q = p$ be a prime for simplicity. There are p choices of x in $\mathbb{F}_p$. For each value x_0, the polynomial $f(x) = x^3 + Ax + B$ gives a value $f(x_0) \in \mathbb{F}_p$. The probability that a random element in $\mathbb{F}_p$ is a perfect square in $\mathbb{F}_p$ is $1/2$ (notice, however, that $f(x_0)$ is not random; this is just a heuristic argument). If $f(x_0)$ is a square modulo p, i.e., if there is a $y_0 \in \mathbb{F}_p$ such that $f(x_0) \equiv y_0^2 \bmod p$, then there are two points $(x_0, \pm y_0)$ in

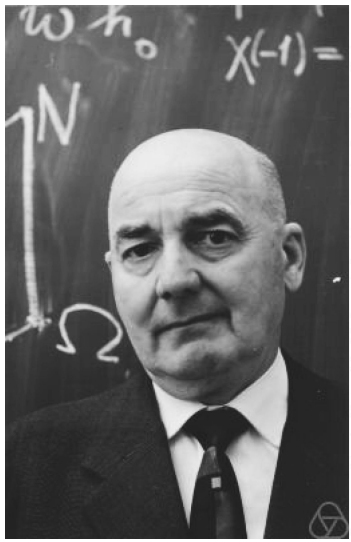


Figure 5. Helmut Hasse (1898-1979).

$\widetilde{E}(\mathbb{F}_p)$. If $f(x_0)$ is not a square modulo p, then there are no points in $\widetilde{E}(\mathbb{F}_p)$ with x-coordinate equal to x_0. Hence:

$$N_p \approx p \cdot \left(\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0 \right) + 1 = p + 1.$$

Notice that we have added 1 in order to account for the point at infinity.

Remark 2.6.13. Suppose that $E/\mathbb{Q}$ is an elliptic curve that has bad reduction at a prime p. How many points does the singular curve $\widetilde{E}$ have over $\mathbb{F}_p$? The reader can find the answer to this question in Exercise 5.7.1.

Example 2.6.14. Let $E/\mathbb{Q}$ be the elliptic curve $y^2 = x^3 + 3$. Its minimal discriminant is $\Delta_E = -3888 = -2^4 \cdot 3^5$. Thus, the only primes of bad reduction are 2 and 3 and $\widetilde{E}/\mathbb{F}_p$ is smooth for all $p \geq 5$. For $p = 5$, there are precisely 6 points on $\widetilde{E}(\mathbb{F}_5)$, namely

$$\widetilde{E}(\mathbb{F}_5) = \{\widetilde{O}, (1,2), (1,3), (2,1), (2,4), (3,0)\},$$

where all the coordinates should be regarded as congruences modulo 5. Thus, $N_5 = 6$, which is in the range given by Hasse's bound:

$$1.5278\ldots = 5 + 1 - 2\sqrt{5} < N_5 < 5 + 1 + 2\sqrt{5} = 10.4721\ldots.$$

Similarly, one can verify that $N_7 = 13$. ■

The connections between the numbers N_p and the group $E(\mathbb{Q})$ are numerous and of great interest. The most surprising relationship is captured by the Birch and Swinnerton-Dyer conjecture (Conjecture 5.2.1) that relates the growth of N_p (as p varies) with the rank of the elliptic curve $E/\mathbb{Q}$. We shall discuss this conjecture in Section 5.2 in more detail. In the next proposition we describe a different connection between N_p and $E(\mathbb{Q})$. We shall use the following notation: if G is an abelian group and $m \geq 2$, then the points of G of order dividing m will be denoted by $G[m]$.

Proposition 2.6.15 (**[Sil86]**, Ch. VII, Prop. 3.1)**.** *Let $E/\mathbb{Q}$ be an elliptic curve, p a prime number and m a natural number not divisible by p. Suppose that $E/\mathbb{Q}$ has good reduction at p. Then the reduction map modulo p,*

$$E(\mathbb{Q})[m] \longrightarrow \widetilde{E}(\mathbb{F}_p),$$

is an injective homomorphism of abelian groups. In particular, the number of elements of $E(\mathbb{Q})[m]$ divides the number of elements of $\widetilde{E}(\mathbb{F}_p)$.

The previous proposition can be very useful when calculating the torsion subgroup of an elliptic curve. Let's see an application:

Example 2.6.16. Let $E/\mathbb{Q}$: $y^2 = x^3 + 3$. In Example 2.6.14 we have seen that $N_5 = 6$ and $N_7 = 13$, and $E/\mathbb{Q}$ has bad reduction only at 2 and 3.

If $q \neq 5, 7$ is a prime number, then $E(\mathbb{Q})[q]$ is trivial. Indeed, Proposition 2.6.15 implies that $|E(\mathbb{Q})[q]|$ divides $N_5 = 6$ and also $N_7 = 13$. Thus, $|E(\mathbb{Q})[q]|$ must divide $\gcd(6, 13) = 1$.

In the case of $q = 5$, we know that $|E(\mathbb{Q})[5]|$ divides $N_7 = 13$. Moreover, by Lagrange's theorem from group theory, if $E(\mathbb{Q})[p]$ is non-trivial, then p divides $|E(\mathbb{Q})[p]|$ (later on we will see that $E(\mathbb{Q})[p]$ is always a subgroup of $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$; see Exercise 3.7.5). Since 5

does not divide 13, it follows that $E(\mathbb{Q})[5]$ must be trivial. Similarly, one can show that $E(\mathbb{Q})[7]$ is trivial, and we conclude that $E(\mathbb{Q})_{\text{torsion}}$ is trivial.

However, notice that $P = (1, 2) \in E(\mathbb{Q})$ is a point on the curve. Since we just proved that E does not have any points of finite order, it follows that P must be a point of *infinite* order, and, hence, we have shown that E has infinitely many rational points: $\pm P, \pm 2P, \pm 3P, \ldots$. In fact, $E(\mathbb{Q}) \cong \mathbb{Z}$ and $(1, 2)$ is a generator of its Mordell-Weil group. ■

In the previous example, the Nagell-Lutz theorem (Theorem 2.5.5) would have yielded the same result, i.e., the torsion is trivial, in an easier way. Indeed, for the curve $E : y^2 = x^3 + 3$, the quantity $4A^3 + 27B^2$ equals 3^5, so the possibilities for $y(P)^2$, where P is a torsion point of order ≥ 3, are 1, 9 or 81 (it is easy to see that there are no 2-torsion points). Therefore, the possibilities for $x(P)^3 = y(P)^2 - 3$ are -2, 6 or 78, respectively. Since $x(P)$ is an integer, we reach a contradiction. In the following example, the Nagell-Lutz theorem would be a lengthier and much more tedious alternative, and Proposition 2.6.15 is much more effective.

Example 2.6.17. Let $E/\mathbb{Q} : y^2 = x^3 + 4249388$. In this case

$$4A^3 + 27B^2 = 2^4 \cdot 3^3 \cdot 11^2 \cdot 13^2 \cdot 17^2 \cdot 19^2 \cdot 23^2.$$

Therefore, $4A^3 + 27B^2$ is divisible by 192 distinct positive squares, which makes it very tedious to use the Nagell-Lutz theorem. The (minimal) discriminant of $E/\mathbb{Q}$ is $\Delta_E = -16(4A^3 + 27B^2)$ and therefore E has good reduction at 5 and 7. Moreover, $B = 4249388 \equiv 3 \bmod 35$ and therefore, by our calculations in Example 2.6.16, $N_5 = 6$ and $N_7 = 13$. Thus, Proposition 2.6.15 and the same argument we used in Ex. 2.6.16 show that the torsion of $E(\mathbb{Q})$ is trivial.

Incidentally, the curve $E/\mathbb{Q} : y^2 = x^3 + 4249388$ has a rational point $P = \left(\frac{25502}{169}, \frac{6090670}{2197}\right)$. Since the torsion of $E(\mathbb{Q})$ is trivial, P must be of infinite order. Another way to see this: since P has rational coordinates that are not integral, the Nagell-Lutz theorem implies that the order of P is infinite. In fact, $E(\mathbb{Q})$ is isomorphic to $\mathbb{Z}$ and it is generated by P. ■

2.7. The rank and the free part of $E(\mathbb{Q})$

In the previous sections we have described simple algorithms that determine the torsion subgroup of $E(\mathbb{Q})$. Recall that the Mordell-Weil theorem (Thm. 2.4.3) says that there is a (non-canonical) isomorphism

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{torsion}} \oplus \mathbb{Z}^{R_E}.$$

Our next goal is to try to find R_E generators of the free part of the Mordell-Weil group. Unfortunately, no algorithm is known that will always yield such free points. We don't even have a way to determine R_E, the rank of the curve, although sometimes we can obtain upper bounds for the rank of a given curve $E/\mathbb{Q}$ (see, for instance, Theorem 2.7.4 below).

Naively, one could hope that if the coefficients of the (minimal) Weierstrass equation for $E/\mathbb{Q}$ are *small*, then the coordinates of the generators of $E(\mathbb{Q})$ should also be *small*, and perhaps a *brute force* computer search would yield these points. However, Bremner and Cassels found the following surprising example: the curve $y^2 = x^3 + 877x$ has rank equal to 1 and the x-coordinate of a generator P is

$$x(P) = (612776083187947368101/788415358606839002 10)^2.$$

However, Serge Lang salvaged this idea and conjectured that for all $\epsilon > 0$ there is a constant C_ϵ such that there is a system of generators $\{P_i : i = 1, \ldots, R_E\}$ of $E(\mathbb{Q})$ with

$$\widehat{h}(P_i) \leq C_\epsilon \cdot |\Delta_E|^{1/2+\epsilon},$$

where $\widehat{h}$ is the canonical height function of $E/\mathbb{Q}$, which we define next. Lang's conjecture says that the size of the coordinates of a generator may grow exponentially with the (minimal) discriminant of a curve $E/\mathbb{Q}$.

Definition 2.7.1. We define the *height* of $\frac{m}{n} \in \mathbb{Q}$, with $\gcd(m, n) = 1$, by

$$h\left(\frac{m}{n}\right) = \log(\max\{|m|, |n|\}).$$

This can be used to define a height on a point $P = (x, y)$ on an elliptic curve $E/\mathbb{Q}$, with $x, y \in \mathbb{Q}$ by

$$H(P) = h(x).$$

Finally, we define the *canonical height* of $P \in E(\mathbb{Q})$ by

$$\widehat{h}(P) = \frac{1}{2} \lim_{N\to\infty} \frac{H(2^N \cdot P)}{4^N}.$$

Note: here $2^N \cdot P$ means multiplication in the curve, using the addition law defined in Section 2.4, i.e., $2 \cdot P = P + P$, $2^2 \cdot P = 2P + 2P$, etc.

Example 2.7.2. Let $E : y^2 = x^3 + 877x$, and let P be a generator of $E(\mathbb{Q})$. Here are some values of $\frac{1}{2} \cdot \frac{H(2^N \cdot P)}{4^N}$:

$$\begin{aligned}
\frac{1}{2} \cdot H(P) &= 47.8645312628\ldots \\
\frac{1}{2} \cdot \frac{H(2 \cdot P)}{4} &= 47.7958126219\ldots \\
\frac{1}{2} \cdot \frac{H(2^2 \cdot P)}{4^2} &= 47.9720107996\ldots \\
\frac{1}{2} \cdot \frac{H(2^3 \cdot P)}{4^3} &= 47.9636902383\ldots \\
\frac{1}{2} \cdot \frac{H(2^4 \cdot P)}{4^4} &= 47.9901607777\ldots \\
\frac{1}{2} \cdot \frac{H(2^5 \cdot P)}{4^5} &= 47.9901600133\ldots \\
\frac{1}{2} \cdot \frac{H(2^6 \cdot P)}{4^6} &= 47.9901569227\ldots \\
\frac{1}{2} \cdot \frac{H(2^7 \cdot P)}{4^7} &= 47.9901419861\ldots \\
\frac{1}{2} \cdot \frac{H(2^8 \cdot P)}{4^8} &= 47.9901807594\ldots.
\end{aligned}$$

The limit is in fact equal to $\widehat{h}(P) = 47.9901859939...$, well below the value $|\Delta_E|^{1/2} = 207,773.12....$ ■

The canonical height enjoys the following properties and, in fact, the canonical height is defined so that it is (essentially) the *only* height that satisfies these properties:

Proposition 2.7.3 (Néron-Tate)**.** *Let $E/\mathbb{Q}$ be an elliptic curve and let $\widehat{h}$ be the canonical height on E.*

(1) *For all $P, Q \in E(\mathbb{Q})$, $\widehat{h}(P+Q)+\widehat{h}(P-Q) = 2\widehat{h}(P)+2\widehat{h}(Q)$. (Note: this is called the parallelogram law.)*

(2) *For all $P \in E(\mathbb{Q})$ and $m \in \mathbb{Z}$, $\widehat{h}(mP) = m^2 \cdot \widehat{h}(P)$. (Note: in particular, the height of mP is much larger than the height of P, for any $m \neq 0, 1$.)*

(3) *Let $P \in E(\mathbb{Q})$. Then $\widehat{h}(P) \geq 0$, and $\widehat{h}(P) = 0$ if and only if P is a torsion point.*

For the proofs of these properties, see **[Sil86]**, Ch. VIII, Thm. 9.3, or **[Mil06]**, Ch. IV, Prop. 4.5 and Thm. 4.7.

As we mentioned at the beginning of this section, we can calculate upper bounds on the rank of a given elliptic curve (see **[Sil86]**, p. 235, exercises 8.1, 8.2). Here is an example:

Theorem 2.7.4 (**[Loz08]**, Prop. 1.1)**.** *Let $E/\mathbb{Q}$ be an elliptic curve given by a Weierstrass equation of the form*

$$E\colon y^2 = x^3 + Ax^2 + Bx, \text{ with } A, B \in \mathbb{Z}.$$

Let R_E be the rank of $E(\mathbb{Q})$. For an integer $N \geq 1$, let $\nu(N)$ be the number of distinct positive prime divisors of N. Then

$$R_E \leq \nu(A^2 - 4B) + \nu(B) - 1.$$

More generally, let $E/\mathbb{Q}$ be any elliptic curve with a non-trivial point of 2-torsion and let a (resp. m) be the number of primes of additive (resp. multiplicative) bad reduction of $E/\mathbb{Q}$. Then

$$R_E \leq m + 2a - 1.$$

Example 2.7.5. Pierre de Fermat proved that $n = 1$ is not a congruent number (see Example 1.1.2) by showing that $x^4 + y^4 = z^2$ has no rational solutions. As an application of the previous theorem, let us show that the curve

$$E_1\colon y^2 = x^3 - x = x(x-1)(x+1)$$

only has the trivial solutions $(0, 0)$, $(\pm 1, 0)$, which are torsion points of order 2. Indeed, the minimal discriminant of E_1 is $\Delta_{E_1} = 64$. Therefore $p = 2$ is the unique prime of bad reduction. Moreover, the reader can check that the reduction at $p = 2$ is multiplicative. Now thanks to Theorem 2.7.4 we conclude that $R_{E_1} = 0$ and E_1 only has torsion points. Finally, using Proposition 2.6.15 or Theorem 2.5.5, we can show that the only torsion points are the three trivial points named above. ■

Example 2.7.6. Let $E/\mathbb{Q}$ be the elliptic curve $y^2 = x(x+1)(x+2)$, which already appeared in Example 1.1.1. Since the Weierstrass equation of E is

$$y^2 = x(x+1)(x+2) = x^3 + 3x^2 + 2x,$$

it follows from Theorem 2.7.4 that the rank R_E satisfies

$$R_E \leq \nu(A^2 - 4B) + \nu(B) - 1 = \nu(1) + \nu(2) - 1 = 0 + 1 - 1 = 0,$$

and therefore the rank is 0. The reader can check that

$$E(\mathbb{Q})_{\text{torsion}} = \{\mathcal{O}, (0,0), (-1,0), (-2,0)\}.$$

Since the rank is zero, the four torsion points on $E/\mathbb{Q}$ are the only rational points on E. ■

Example 2.7.7. Let $E : y^2 = x^3 + 2308x^2 + 665858x$. The primes 2 and 577 are the only prime divisors of (both) B and $A^2 - 4B$. Thus,

$$R_E \leq \nu(A^2 - 4B) + \nu(B) - 1 = 2 + 2 - 1 = 3.$$

The points $P_1 = (-1681, 25543)$, $P_2 = (-338, 26)$, and $P_3 = (577/16, 332929/64)$ are of infinite order and the subgroup of $E(\mathbb{Q})$ generated by P_1, P_2 and P_3 is isomorphic to $\mathbb{Z}^3$. Therefore, the rank of E is equal to 3. ■

2.8. Linear independence of rational points

Let $E/\mathbb{Q}$ be the curve defined in Example 2.7.7. We claimed that the subgroup generated by the points $P_1 = (-1681, 25543)$, $P_2 = (-338, 26)$, and $P_3 = (577/16, 332929/64)$ is isomorphic to $\mathbb{Z}^3$. But how can we show that? In particular, why is P_3 not a linear combination of P_1 and P_2? In other words, are there integers n_1 and n_2 such that $P_3 = n_1P_1 + n_2P_2$? In fact, $E/\mathbb{Q}$ has a rational torsion point $T = (0,0)$ of order 2, so could some combination of P_1, P_2 and P_3 equal T? This example motivates the need for a notion of linear dependence and independence of points over $\mathbb{Z}$.

Definition 2.8.1. Let $E/\mathbb{Q}$ be an elliptic curve. We say that the rational points $P_1, \ldots, P_m \in E(\mathbb{Q})$ are *linearly dependent over* $\mathbb{Z}$ if there are integers $n_1, \ldots, n_m \in \mathbb{Z}$ such that

$$n_1P_1 + n_2P_2 + \cdots + n_mP_m = T,$$

where T is a torsion point. Otherwise, if no such relation exists, we say that the points are *linearly independent* over $\mathbb{Z}$.

Example 2.8.2. Let $E/\mathbb{Q} : y^2 = x^3 + x^2 - 25x + 39$ and let

$$P_1 = \left(\frac{61}{4}, -\frac{469}{8}\right),\ P_2 = \left(-\frac{335}{81}, -\frac{6868}{729}\right),\ P_3 = (21, 96).$$

The points P_1, P_2 and P_3 are rational points on E and linearly dependent over $\mathbb{Z}$ because

$$-3P_1 - 2P_2 + 6P_3 = \mathcal{O}.$$

■

Example 2.8.3. Let $E/\mathbb{Q} : y^2 + y = x^3 - x^2 - 26790x + 1696662$ and put

$$\begin{aligned} P_1 &= \left(\frac{59584}{625}, \frac{71573}{15625}\right), \\ P_2 &= \left(\frac{101307506181}{210337009}, \frac{30548385002405573}{3050517641527}\right). \end{aligned}$$

The points P_1 and P_2 are rational points on E, and they are linearly dependent over $\mathbb{Z}$ because

$$-3P_1 + 2P_2 = (133, -685),$$

and $(133, -685)$ is a torsion point of order 5. ■

Now that we have defined linear independence over $\mathbb{Z}$, we need a method to prove that a number of points are linearly independent. The existence of the Néron-Tate pairing provides a way to prove independence.

Definition 2.8.4. The *Néron-Tate pairing* attached to an elliptic curve is defined by

$$\langle \cdot, \cdot \rangle : E(\mathbb{Q}) \times E(\mathbb{Q}) \to \mathbb{R}, \quad \langle P, Q \rangle = \widehat{h}(P + Q) - \widehat{h}(P) - \widehat{h}(Q),$$

where $\widehat{h}$ is the canonical height on E. Let $P_1, P_2, \ldots, P_r$ be r rational points on $E(\mathbb{Q})$. The *elliptic height matrix* associated to $\{P_i\}_{i=1}^r$ is

$$\mathcal{H} = \mathcal{H}(\{P_i\}_{i=1}^r) := (\langle P_i, P_j \rangle)_{1 \le i \le r,\ 1 \le j \le r}.$$

The determinant of $\mathcal{H}$ is called the *elliptic regulator* of the set of points $\{P_i\}_{i=1}^r$. If $\{P_i\}_{i=1}^r$ is a complete set of generators of the free

part of $E(\mathbb{Q})$, then the determinant of $\mathcal{H}(\{P_i\}_{i=1}^r)$ is called the *elliptic regulator of* $E/\mathbb{Q}$.

Theorem 2.8.5. *Let $E/\mathbb{Q}$ be an elliptic curve. Then the Néron-Tate pairing $\langle\cdot,\cdot\rangle$ associated to E is a non-degenerate symmetric bilinear form on $E(\mathbb{Q})/E(\mathbb{Q})_{torsion}$, i.e.,*

(1) *For all $P, Q \in E(\mathbb{Q})$, $\langle P, Q\rangle = \langle Q, P\rangle$.*

(2) *For all $P, Q, R \in E(\mathbb{Q})$ and all $m, n \in \mathbb{Z}$,*

$$\langle P, mQ + nR\rangle = m\langle P, Q\rangle + n\langle P, R\rangle.$$

(3) *Suppose $P \in E(\mathbb{Q})$ and $\langle P, Q\rangle = 0$ for all $Q \in E(\mathbb{Q})$. Then $P \in E(\mathbb{Q})_{torsion}$. In particular, P is a torsion point if and only if $\langle P, P\rangle = 0$.*

The properties of the Néron-Tate pairing follow from those of the canonical height in Proposition 2.7.3 (see Exercise 2.12.12). Theorem 2.8.5 has the following important corollary:

Corollary 2.8.6. *Let $E/\mathbb{Q}$ be an elliptic curve and let $P_1, P_2, \ldots, P_r \in E(\mathbb{Q})$ be rational points. Let $\mathcal{H}$ be the elliptic height matrix associated to $\{P_i\}_{i=1}^r$. Then:*

(1) *Suppose $\det(\mathcal{H}) = 0$ and $u = (n_1, \ldots, n_r) \in \mathrm{Ker}(\mathcal{H})$, with $n_i \in \mathbb{Z}$. Then the points $\{P_i\}_{i=1}^r$ are linearly dependent and $\sum_{k=1}^r n_k P_k = T$, where T is a torsion point on $E(\mathbb{Q})$.*

(2) *If $\det(\mathcal{H}) \neq 0$, then the points $\{P_i\}_{i=1}^r$ are linearly independent and the rank of $E(\mathbb{Q})$ is $\geq r$.*

Here is an example of how the Néron-Tate pairing is used in practice:

Example 2.8.7. Let $E/\mathbb{Q}$ be the elliptic curve $y^2 = x^3 + 2308x^2 + 665858x$. Put

$$\begin{aligned} P &= (-1681, 25543), \quad Q = (-338, 26), \quad \text{and} \\ R &= \left(\frac{332929}{36}, -\frac{215405063}{216}\right). \end{aligned}$$

Are P, Q and R independent? In order to find out, we find the elliptic height matrix associated to $\{P, Q, R\}$, using PARI or Sage:

$$\begin{aligned}\mathcal{H} &= \begin{pmatrix} \langle P,P\rangle & \langle Q,P\rangle & \langle R,P\rangle \\ \langle P,Q\rangle & \langle Q,Q\rangle & \langle R,Q\rangle \\ \langle P,R\rangle & \langle Q,R\rangle & \langle R,R\rangle \end{pmatrix} \\ &= \begin{pmatrix} 7.397\ldots & -3.601\ldots & 3.795\ldots \\ -3.601\ldots & 6.263\ldots & 2.661\ldots \\ 3.795\ldots & 2.661\ldots & 6.457\ldots \end{pmatrix}.\end{aligned}$$

The determinant of $\mathcal{H}$ seems to be *very* close to 0 (PARI returns $3.368 \cdot 10^{-27}$). Hence Cor. 2.8.6 suggests that P, Q and R are not independent. If we find the (approximate) kernel of $\mathcal{H}$ with PARI, we discover that the (column) vector $(1, 1, -1)$ is approximately in the kernel, and therefore, $P + Q - R$ may be a torsion point. Indeed, the point $P+Q-R = (0, 0)$ is a torsion point of order 2 on $E(\mathbb{Q})$. Hence, P, Q and R are linearly *dependent* over $\mathbb{Z}$.

Instead, let $P_1 = (-1681, 25543)$, $P_2 = (-338, 26)$, a third point $P_3 = (577/16, 332929/64)$ and let $\mathcal{H}'$ be the elliptic height matrix associated to $\{P_i\}_{i=1}^3$. Then $\det(\mathcal{H}') = 101.87727\ldots$ is non-zero and, therefore, $\{P_i\}_{i=1}^3$ are linearly independent and the rank of $E/\mathbb{Q}$ is at least 3. ■

2.9. Descent and the weak Mordell-Weil theorem

In the previous sections we have seen methods to calculate the torsion subgroup of an elliptic curve $E/\mathbb{Q}$, and also methods to check if a collection of points are independent modulo torsion. However, we have not discussed any method to find points of infinite order. In this section, we briefly explain the *method of descent*, which facilitates the search for generators of the free part of $E(\mathbb{Q})$. Unfortunately, the method of descent is not always successful! We will try to measure the failure of the method in the following section. The method of descent (as explained here) is mostly due to Cassels. For a more detailed treatment, see [**Was08**] or [**Sil86**]. A more general descent algorithm was laid out by Birch and Swinnerton-Dyer in [**BSD63**].

The current implementation of the algorithm is more fully explained in Cremona's book **[Cre97]**.

Let $E/\mathbb{Q}$ be a curve given by $y^2 = x^3 + Ax + B$, with $A, B \in \mathbb{Z}$. The most general case of the method of descent is quite involved, so we will concentrate on a particular case where the calculations are much easier: we will assume that $E(\mathbb{Q})$ has 4 distinct rational points of 2-torsion (including $\mathcal{O}$). As we saw before (Theorem 2.5.5, or Exercise 2.12.6), a point $P = (x, y) \in E(\mathbb{Q})$ is of 2-torsion if and only if $y = 0$ and $x^3 + Ax + B = 0$ (or $P = \mathcal{O}$). Thus, if $E(\mathbb{Q})$ has 4 distinct rational points of order 2, that means that $x^3 + Ax + B$ has three (integral) roots and it factors completely over $\mathbb{Z}$:

$$x^3 + Ax + B = (x - e_1)(x - e_2)(x - e_3)$$

with $e_i \in \mathbb{Z}$. Since $x^3 + Ax + B$ does not have an x^2 term, we conclude that $e_1 + e_2 + e_3 = 0$.

Suppose, then, that $E : y^2 = (x - e_1)(x - e_2)(x - e_3)$, where the roots satisfy $e_i \in \mathbb{Z}$ and $e_1 + e_2 + e_3 = 0$. We would like to find a solution $(x_0, y_0) \in E$ with $x_0, y_0 \in \mathbb{Q}$, i.e.,

$$y_0^2 = (x_0 - e_1)(x_0 - e_2)(x_0 - e_3).$$

Thus, each term $(x_0 - e_i)$ must be *almost* a square, and we can make this precise by writing

$$(x_0 - e_1) = au^2,\ (x_0 - e_2) = bv^2,\ (x_0 - e_3) = cw^2,\ y_0^2 = abc(uvw)^2,$$

where $a, b, c, u, v, w \in \mathbb{Q}$, the numbers $a, b, c \in \mathbb{Q}$ are square-free, and abc is a square (in $\mathbb{Q}$).

Example 2.9.1. Let

$$E : y^2 = x^3 - 556x + 3120 = (x - 6)(x - 20)(x + 26)$$

so that $e_1 = 6$, $e_2 = 20$ and $e_3 = -26$. The point $(x_0, y_0) = (\frac{164184}{289}, \frac{66469980}{4913})$ is rational and on E. We can write

$$x_0 - e_1 = \frac{164184}{289} - 6 = 2 \cdot \left(\frac{285}{17}\right)^2$$

and, similarly, $x_0 - e_2 = (\frac{398}{17})^2$ and $x_0 - e_3 = 2 \cdot (\frac{293}{17})^2$. Thus, following the notation of the preceeding paragraphs

$$a = 2,\ b = 1,\ c = 2,\ u = \frac{285}{17},\ v = \frac{398}{17},\ w = \frac{293}{17}.$$

Notice that abc is a square and $y_0^2 = (\frac{66469980}{4913})^2 = abc(uvw)^2$. ■

Example 2.9.2. Let $E : y^2 = x^3 - 556x + 3120$ as before, with $e_1 = 6$, $e_2 = 20$ and $e_3 = -26$. Let $P = (-8, 84)$, $Q = (24, 60)$ and $S = P+Q = (-\frac{247}{16}, -\frac{5733}{64})$. The points P, Q and S are in $E(\mathbb{Q})$. We would like to calculate the aforementioned numbers a, b, c for each of the points P, Q and S. For instance

$$\begin{aligned} x(P) - e_1 &= -8 - 6 = -14 = -14 \cdot 1^2, \\ x(P) - e_2 &= -7 \cdot 4^2, \text{ and } x(P) - e_3 = 2 \cdot 3^2. \end{aligned}$$

Thus, $a_P = -14$, $b_P = -7$ and $c_P = 2$. Similarly, we calculate

$$\begin{aligned} x(Q) - 6 &= 2 \cdot 3^2, \quad x(Q) - 20 = 2^2, \quad x(Q) + 26 = 2 \cdot 5^2, \\ x(S) - 6 &= -7 \cdot \left(\frac{7}{4}\right)^2, \\ x(S) - 20 &= -7 \cdot \left(\frac{9}{4}\right)^2, \; x(S) + 26 = \left(\frac{13}{4}\right)^2. \end{aligned}$$

Thus $a_Q = 2$, $b_Q = 1$, $c_Q = 2$, and $a_S = -7$, $b_S = -7$, $c_S = 1$. Notice the following interesting fact:

$$a_P \cdot a_Q = -28 = -7 \cdot 2^2, \quad b_P \cdot b_Q = -7, \quad c_P \cdot c_Q = 4.$$

Therefore, the square-free part of $a_P \cdot a_Q$ equals $a_S = a_{P+Q} = -7$. And similarly, the square-free parts of $b_P \cdot b_Q$ and $c_P \cdot c_Q$ equal $b_S = -7$ and $c_S = 1$, respectively. Also, the reader can check that $a_{2P} = b_{2P} = c_{2P} = 1$ and $a_{2Q} = b_{2Q} = c_{2Q} = 1$. ■

The previous example points to the fact that there may be a homomorphism between points on $E(\mathbb{Q})$ and triples (a, b, c) of rational numbers modulo squares, or square-free parts of rational numbers; formally, we are talking about $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$. Here, the group $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ is the multiplicative group of non-zero rational numbers, with the extra relation that two non-zero rational numbers are equivalent if their square-free parts are equal (or, equivalently, if their quotient is a perfect square). For instance, 3 and $\frac{12}{25}$ represent the same element of $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ because $\frac{12}{25} = 3 \cdot \left(\frac{2}{5}\right)^2$. The following theorem constructs such a homomorphism. Here we have adapted the proof that appears in **[Was08]**, Theorem 8.14.

Theorem 2.9.3. *Let $E/\mathbb{Q}$ be an elliptic curve*

$$y^2 = x^3 + Ax + B = (x - e_1)(x - e_2)(x - e_3)$$

with distinct $e_1, e_2, e_3 \in \mathbb{Z}$ and $e_1 + e_2 + e_3 = 0$. There is a homomorphism of groups

$$\delta : E(\mathbb{Q}) \to \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$$

defined for $P = (x_0, y_0)$ by

$$\delta(P) = \begin{cases} (1,1,1) & \text{if } P = \mathcal{O}; \\ (x_0 - e_1, x_0 - e_2, x_0 - e_3) & \text{if } y_0 \neq 0; \\ ((e_1 - e_2)(e_1 - e_3), e_1 - e_2, e_1 - e_3) & \text{if } P = (e_1, 0); \\ (e_2 - e_1, (e_2 - e_1)(e_2 - e_3), e_2 - e_3) & \text{if } P = (e_2, 0); \\ (e_3 - e_1, e_3 - e_2, (e_3 - e_1)(e_3 - e_2)) & \text{if } P = (e_3, 0). \end{cases}$$

If $\delta(P) = (\delta_1, \delta_2, \delta_3)$, then $\delta_1 \cdot \delta_2 \cdot \delta_3 = 1$ in $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$. Moreover, the kernel of δ is precisely $2E(\mathbb{Q})$; i.e., if $\delta(Q) = (1,1,1)$, then $Q = 2P$ for some $P \in E(\mathbb{Q})$.

Proof. Let δ be the function defined in the statement of the theorem. Let us show that δ is a homomorphism of (abelian) groups; i.e., we want to show that $\delta(P) \cdot \delta(Q) = \delta(P + Q)$. Notice first of all that $\delta(P) = \delta(x_0, y_0) = \delta(x_0, -y_0) = \delta(-P)$, because the definition of δ does not depend on the sign of the y coordinate of P (in fact, it only depends on whether $y(P) = 0$). Thus, it suffices to prove that $\delta(P) \cdot \delta(Q) = \delta(-(P + Q))$ for all $P, Q \in E(\mathbb{Q})$.

Let $P = (x_0, y_0)$, $Q = (x_1, y_1)$ and $R = -(P + Q) = (x_2, y_2)$, and let us assume, for simplicity, that $y_i \neq 0$ for $i = 1, 2, 3$. By the definition of the addition rule on an elliptic curve (see Figure 1), the points P, Q and R are collinear. Let $\mathcal{L} = \overline{PQ}$ be the line that goes through all three points, and suppose it has equation $\mathcal{L} : y = ax + b$. Therefore, if we substitute y in the equation of $E/\mathbb{Q}$, we obtain a polynomial

$$p(x) = (ax + b)^2 - (x - e_1)(x - e_2)(x - e_3).$$

The polynomial $p(x)$ is cubic, its leading term is -1, and it has precisely three rational roots, namely x_0, x_1 and x_2. Hence, it factors:

$$p(x) = (ax+b)^2 - (x-e_1)(x-e_2)(x-e_3) = -(x-x_0)(x-x_1)(x-x_2).$$

If we evaluate $p(x)$ at $x = e_i$, we obtain

$$p(e_i) = (ae_i + b)^2 = -(e_i - x_0)(e_i - x_1)(e_i - x_2)$$

or, equivalently, $(x_0 - e_i)(x_1 - e_i)(x_2 - e_i) = (ae_i + b)^2$. Thus, the product $\delta(P) \cdot \delta(Q) \cdot \delta(R)$ equals

$$\begin{aligned}
\delta(P) \cdot \delta(Q) \cdot \delta(R) &= (x_0 - e_1, x_0 - e_2, x_0 - e_3) \\
&\quad \cdot (x_1 - e_1, x_1 - e_2, x_1 - e_3) \\
&\quad \cdot (x_2 - e_1, x_2 - e_2, x_2 - e_3) \\
&= ((x_0 - e_1)(x_1 - e_1)(x_2 - e_1), \\
&\quad (x_0 - e_2)(x_1 - e_2)(x_2 - e_2), \\
&\quad (x_0 - e_3)(x_1 - e_3)(x_2 - e_3)) \\
&= ((ae_1 + b)^2, (ae_2 + b)^2, (ae_3 + b)^2) \\
&= (1, 1, 1) \in (\mathbb{Q}^\times/(\mathbb{Q}^\times)^2)^3.
\end{aligned}$$

Hence, $\delta(P) \cdot \delta(Q) \cdot \delta(R) = 1$. If we multiply both sides by $\delta(R)$ and notice that $a^2 = 1$ for any $a \in \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$, we conclude that

$$\delta(P) \cdot \delta(Q) = \delta(R) = \delta(-(P + Q)) = \delta(P + Q),$$

as desired. In order to completely prove that δ is a homomorphism, we would need to check the cases when P, Q or R is one of the points $(e_i, 0)$ or $\mathcal{O}$, but we leave those special cases for the reader to check (Exercise 2.12.15).

If $\delta(P) = (\delta_1, \delta_2, \delta_3)$, then it follows directly from the definition of δ that $\delta_1 \cdot \delta_2 \cdot \delta_3 = 1$ in $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$. Indeed, this is clear for $P = \mathcal{O}$ or $P = (e_i, 0)$, and if $P = (x_0, y_0)$ with $y_0 \neq 0$, then $(x_0 - e_1)(x_0 - e_2)(x_0 - e_3) = y_0^2$, which is a square, and is therefore trivial in $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$.

Next, let us show that the kernel of δ is $2E(\mathbb{Q})$. Clearly, $2E(\mathbb{Q})$ is in the kernel of δ, because δ is a homomorphism with image in $(\mathbb{Q}^\times/(\mathbb{Q}^\times)^2)^3$, as we just proved. Indeed, if $P \in E(\mathbb{Q})$, then

$$\delta(2P) = \delta(P) \cdot \delta(P) = \delta(P)^2 = (\delta_1^2, \delta_2^2, \delta_3^2) = (1, 1, 1),$$

because squares are trivial in $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$.

Now let us show the reverse inclusion, i.e., that the kernel of δ is contained in $2E(\mathbb{Q})$. Let $Q = (x_1, y_1) \in E(\mathbb{Q})$ such that $\delta(Q) = (1, 1, 1)$. We want to find $P = (x_0, y_0)$ such that $2P = Q$. Notice that

it is enough to show that $x(2P) = x_1$, because $2P$ is a point on $E(\mathbb{Q})$ and if $x(2P) = x(Q)$, then $Q = 2(\pm P)$. Hence, our goal will be to construct $(x_0, y_0) \in E(\mathbb{Q})$ such that

$$x(2P) = \frac{x_0^4 - 2Ax_0^2 - 8Bx_0 + A^2}{4y_0^2} = x_1.$$

The formula for $x(2P)$ above is given in Exercise 2.12.16.

Once again, for simplicity, let us assume $y(Q) = y_1 \neq 0$ and, as stated above, we assume $\delta(Q) = (1, 1, 1)$. Hence, $x_1 - e_i$ is a square in $\mathbb{Q}$ for $i = 1, 2, 3$. Let us write

$$x_1 - e_i = t_i^2, \quad \text{for some } t_i \in \mathbb{Q}^\times. \tag{2.6}$$

We define a new auxiliary polynomial $p(x)$ by

$$t_1 \frac{(x-e_2)(x-e_3)}{(e_1-e_2)(e_1-e_3)} + t_2 \frac{(x-e_1)(x-e_3)}{(e_2-e_1)(e_2-e_3)} + t_3 \frac{(x-e_1)(x-e_2)}{(e_3-e_1)(e_3-e_2)}.$$

The polynomial $p(x)$ is an interpolating polynomial (or Lagrange polynomial) which was defined so that $p(e_i) = t_i$. Notice that $p(x)$ is a quadratic polynomial, say $p(x) = a + bx + cx^2$. Also define another polynomial $q(x) = x_1 - x - p(x)^2$ and notice that

$$q(e_i) = x_1 - e_i - p(e_i)^2 = x_1 - e_i - t_i^2 = 0$$

from the definition of t_i in Eq. (2.6). Since $q(e_i) = 0$, it follows that $(x - e_i)$ divides $q(x)$ for $i = 1, 2, 3$. Thus, $(x - e_1)(x - e_2)(x - e_3) = x^3 + Ax + B$ divides $q(x)$. In other words, $q(x) \equiv 0 \bmod x^3 + Ax + B$. Since $q(x) = x_1 - x - p(x)^2$, we can also write

$$x_1 - x \equiv p(x)^2 \equiv (a + bx + cx^2)^2 \bmod (x^3 + Ax + B).$$

We shall expand the square on the right-hand side, modulo $f(x) = x^3 + Ax + B$. Notice that $x^3 \equiv -Ax - B$, and $x^4 \equiv -Ax^2 - Bx$ modulo $f(x)$:

$$\begin{aligned}
x_1 - x &\equiv p(x)^2 \equiv (a + bx + cx^2)^2 \\
&\equiv c^2x^4 + 2bcx^3 + (2ac + b^2)x^2 + 2abx + a^2 \\
&\equiv c^2(-Ax^2 - Bx) + 2bc(-Ax - B) \\
&\quad + (2ac + b^2)x^2 + 2abx + a^2 \\
&\equiv (2ac + b^2 - Ac^2)x^2 \\
&\quad + (2ab - Bc^2 - 2Abc)x + (a^2 - 2bcB),
\end{aligned}$$

where all the congruences are modulo $f(x) = x^3 + Ax + B$. The congruences in the previous equation say that a polynomial of degree 1, call it $g(x) = x_1 - x$, is congruent to a polynomial of degree ≤ 2, call the last line $h(x)$, modulo a polynomial of degree 3, namely $f(x)$. Then $h(x) - g(x)$ is a polynomial of degree ≤ 2, divisible by a polynomial of degree 3. This implies that $h(x) - g(x)$ must be zero and $h(x) = g(x)$, i.e.,

$$x_1 - x = (2ac + b^2 - Ac^2)x^2 + (2ab - Bc^2 - 2Abc)x + (a^2 - 2bcB).$$

If we match coefficients, we obtain the following equalities:

$$2ac + b^2 - Ac^2 = 0, \tag{2.7}$$

$$2ab - Bc^2 - 2Abc = -1, \tag{2.8}$$

$$a^2 - 2bcB = x_1. \tag{2.9}$$

If $c = 0$, then $b = 0$ by Eq. (2.7); therefore, $p(x) = a + bx + cx^2 = a$ is a constant function, and so $t_1 = t_2 = t_3$. By Eq. (2.6), it follows that $e_1 = e_2 = e_3$, which is a contradiction with our assumptions. Hence, c must be non-zero. We multiply Eq. (2.8) by $\frac{1}{c^2}$ and Eq. (2.7) by $\frac{b}{c^3}$ to obtain

$$\frac{2ab}{c^2} - B - \frac{2Ab}{c} = -\frac{1}{c^2}, \tag{2.10}$$

$$\frac{2ab}{c^2} + \frac{b^3}{c^3} - \frac{Ab}{c} = 0. \tag{2.11}$$

We subtract Eq. (2.10) from Eq. (2.11) to get:

$$\left(\frac{b}{c}\right)^3 + A\left(\frac{b}{c}\right) + B = \left(\frac{1}{c}\right)^2.$$

Hence, the point $P = (x_0, y_0) = (\frac{b}{c}, \frac{1}{c})$ is a rational point on $E(\mathbb{Q})$. It remains to show that $x(2P) = x(Q)$. From Eq. (2.11) we deduce that

$$a = \frac{\frac{Ab}{c} - \frac{b^3}{c^3}}{\frac{2b}{c^2}} = \frac{A - \left(\frac{b}{c}\right)^2}{2 \cdot \frac{1}{c}} = \frac{A - x_0^2}{2y_0},$$

and, therefore, substituting a in Eq. (2.9) yields

$$\begin{aligned} x(Q) = x_1 = a^2 - 2bcB &= \left(\frac{A - x_0^2}{2y_0}\right)^2 - 2bcB \\ &= \frac{(A^2 - 2Ax_0^2 + x_0^4) - (2bcB)(4y_0^2)}{4y_0^2} \\ &= \frac{(A^2 - 2Ax_0^2 + x_0^4) - (2bcB)(\frac{4}{c^2})}{4y_0^2} \\ &= \frac{(A^2 - 2Ax_0^2 + x_0^4) - 8Bx_0}{4y_0^2} \\ &= \frac{x_0^4 - 2Ax_0^2 - 8Bx_0 + A^2}{4y_0^2} = x(2P) \end{aligned}$$

as desired. In order to complete the proof of the fact that the kernel of δ is $2E(\mathbb{Q})$, we would need to consider the case when $y(Q) = y_1 = 0$, but we leave this special case to the reader (Exercise 2.12.18). ■

Thus, the previous proposition shows that there is a homomorphism $\delta : E(\mathbb{Q}) \to (\mathbb{Q}^\times/(\mathbb{Q}^\times)^2)^3$ with kernel equal to $2E(\mathbb{Q})$. In fact, the theorem shows that there is a homomorphism from $E(\mathbb{Q})$ into

$$\Gamma = \{(\delta_1, \delta_2, \delta_3) \in (\mathbb{Q}^\times/(\mathbb{Q}^\times)^2)^3 : \delta_1 \cdot \delta_2 \cdot \delta_3 = 1 \in \mathbb{Q}^\times/(\mathbb{Q}^\times)^2\}.$$

Hence, δ induces an injection

$$E(\mathbb{Q})/2E(\mathbb{Q}) \hookrightarrow \Gamma \subset (\mathbb{Q}^\times/(\mathbb{Q}^\times)^2)^3.$$

The groups $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ and Γ are infinite, so such an injection does not tell us much about the size of $E(\mathbb{Q})/2E(\mathbb{Q})$. However, the image of $E(\mathbb{Q})/2E(\mathbb{Q})$ is much smaller than Γ.

Example 2.9.4. Let $E : y^2 = x^3 - 556x + 3120$ as in Example 2.9.2. It turns out that $E(\mathbb{Q}) \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}^2$. The generators of the torsion part are $T_1 = (6, 0)$ and $T_2 = (20, 0)$, and the generators of the free part are $P = (-8, 84)$ and $Q = (24, 60)$. The image of the map δ in this case is, therefore, generated by the images of T_1, T_2, P and Q.

$$\begin{aligned} \delta(T_1) &= (-7, -14, 2), \quad \delta(T_2) = (14, 161, 46), \\ \delta(P) &= (-14, -7, 2), \quad \delta(Q) = (2, 1, 2). \end{aligned}$$

Thus, the image of δ is formed by the 16 elements that one obtains by multiplying out $\delta(T_1)$, $\delta(T_2)$, $\delta(P)$ and $\delta(Q)$, in all possible ways. Thus, $\delta(E(\mathbb{Q})/2E(\mathbb{Q}))$ is the group:

$$\{(1,1,1),\ (-7,-14,2),\ (14,161,46),\ (-2,-46,23),$$
$$(-14,-7,2),\ (2,2,1),\ (-1,-23,23),\ (7,322,46),$$
$$(2,1,2),\ (-14,-14,1),\ (7,161,23),\ (-1,-46,46),$$
$$(-7,-7,1),\ (1,2,2),\ (-2,-23,46),\ (14,322,23)\}.$$

(Exercise: Check that the elements listed above form a group under multiplication.) We see that the only primes that appear in the factorization of the coordinates of elements in the image of δ are: $2, 7$ and 23. Therefore, the coordinates of δ are not just in $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ but in a much smaller subgroup of 16 elements:

$$\Gamma' = \{\pm 1,\ \pm 2,\ \pm 7,\ \pm 23,\ \pm 14,\ \pm 46,\ \pm 161,\ \pm 322\} \subset \mathbb{Q}^\times/(\mathbb{Q}^\times)^2.$$

And the image of $E(\mathbb{Q})/2E(\mathbb{Q})$ embeds into

$$\begin{aligned}\Gamma_\Delta &= \{(\delta_1,\delta_2,\delta_3) \in \Gamma' \times \Gamma' \times \Gamma' : \delta_1 \cdot \delta_2 \cdot \delta_3 = 1 \in \mathbb{Q}^\times/(\mathbb{Q}^\times)^2\} \\ &\subset \Gamma' \times \Gamma' \times \Gamma'.\end{aligned}$$

Since Γ' has 16 elements and $E(\mathbb{Q})/2E(\mathbb{Q})$ embeds into $(\Gamma')^3$, we conclude that $E(\mathbb{Q})/2E(\mathbb{Q})$ has at most $(16)^3 = 2^{12}$ elements. In fact, Γ_Δ has only 16^2 elements, so $E(\mathbb{Q})/2E(\mathbb{Q})$ has at most 2^8 elements. Notice also the following interesting "coincidence": the prime divisors that appear in Γ_Δ coincide with the prime divisors of the discriminant of E, which is $\Delta_E = 6795034624 = 2^{18}\cdot 7^2 \cdot 23^2$. In the next proposition we explain that, in fact, this is always the case. ■

Proposition 2.9.5. *Let $E : y^2 = (x-e_1)(x-e_2)(x-e_3)$, with $e_i \in \mathbb{Z}$. Let $P = (x_0, y_0) \in E(\mathbb{Q})$ and write*

$$(x_0 - e_1) = au^2,\ (x_0 - e_2) = bv^2,\ (x_0 - e_3) = cw^2,\ y_0^2 = abc(uvw)^2,$$

where $a, b, c, u, v, w \in \mathbb{Q}$, the numbers $a, b, c \in \mathbb{Z}$ are square-free, and abc is a square (in $\mathbb{Z}$). Then, if p divides $a \cdot b \cdot c$, then p also divides the quantity $\Delta = (e_1 - e_2)(e_2 - e_3)(e_1 - e_3)$.

Note: the discriminant of E equals $\Delta_E = 16(e_1 - e_2)^2(e_2 - e_3)^2(e_1 - e_3)^2$. So a prime p divides Δ if and only if p divides Δ_E. If $p > 2$, then this is clear (see Exercise 2.12.19 for $p = 2$).

Proof. Suppose a prime p divides abc. Then p divides a, b or c. Let us assume that $p \mid a$ (the same argument works if p divides b or c). Let p^k be the exact power of p that appears in the factorization of the rational number $x_0 - e_1 = au^2$. Notice that k may be positive or negative, depending on whether p divides the numerator or denominator of au^2. Notice, however, that k must be odd, because $p \mid a$, and a is square-free.

Suppose first that $k < 0$, i.e., $p^{|k|}$ is the exact power of p that divides the denominator of $x_0 - e_1$. Since $e_i \in \mathbb{Z}$, it follows that $p^{|k|}$ must divide the denominator of x_0 too, and therefore $p^{|k|}$ is the exact power that divides the denominators of $x_0 - e_2$ and $x_0 - e_3$ as well. Hence, $p^{3|k|}$ is the exact power of p dividing the denominator of $y_0^2 = \prod(x_0 - e_i)$, but this is impossible because y_0^2 is a square and $3|k|$ is odd. Thus, k must be positive.

If $k > 0$ and p divides $x_0 - e_1$, then the denominator of x_0 is not divisible by p, so it makes sense to consider $x_0 \bmod p$, and $x_0 \equiv e_1 \bmod p$. Similarly, the denominators of $x_0 - e_2$ and $x_0 - e_3$ are not divisible by p and

$$bv^2 \equiv x_0 - e_2 \equiv e_1 - e_2, \quad \text{and} \quad cw^2 \equiv x_0 - e_3 \equiv e_1 - e_3 \bmod p.$$

Since $y_0^2 = abc(uvw)^2$ and p divides a, then p must also divide one of b or c. Let's suppose it also divides b. Then $0 \equiv bv^2 \equiv x_0 - e_2 \equiv e_1 - e_2 \bmod p$ and $\Delta = (e_1 - e_2)(e_2 - e_3)(e_1 - e_3) \equiv 0 \bmod p$, as claimed. ∎

The definition of the map δ and the previous proposition yield the following immediate corollary:

Corollary 2.9.6. *With notation as in the previous Theorem and Proposition, define a subgroup Γ' of $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ by*

$$\Gamma' = \{n \in \mathbb{Z} : 0 \neq n \text{ is square-free and if } p \mid n, \text{ then } p \mid \Delta\}/(\mathbb{Z}^\times)^2.$$

Then, δ induces an injection of $E(\mathbb{Q})/2E(\mathbb{Q})$ into

$$\begin{aligned} \Gamma_\Delta &= \{(\delta_1, \delta_2, \delta_3) \in \Gamma' \times \Gamma' \times \Gamma' : \delta_1 \cdot \delta_2 \cdot \delta_3 = 1 \in \mathbb{Q}^\times/(\mathbb{Q}^\times)^2\} \\ &\subset \Gamma' \times \Gamma' \times \Gamma'. \end{aligned}$$

We are ready to prove the weak Mordell-Weil theorem (Thm. 2.4.5), at least in our restricted case:

Corollary 2.9.7 (Weak Mordell-Weil theorem)**.** *Let $E : y^2 = (x - e_1)(x-e_2)(x-e_3)$ be an elliptic curve, with $e_i \in \mathbb{Z}$. Then $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite.*

Proof. By Cor. 2.9.6, $E(\mathbb{Q})/2E(\mathbb{Q})$ injects into $\Gamma_\Delta \subset \Gamma' \times \Gamma' \times \Gamma'$. Since Γ' is finite, $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite as well. ■

2.10. Homogeneous spaces

In this section we want to make the weak Mordell-Weil theorem explicit, i.e., we want:

- explicit bounds on the size of $E(\mathbb{Q})/2E(\mathbb{Q})$, and
- a method to find generators of $E(\mathbb{Q})/2E(\mathbb{Q})$ (see Exercise 2.12.25, though).

Before we discuss bounds, we need to understand the structure of the quotient $E(\mathbb{Q})/2E(\mathbb{Q})$. Remember that, from the Mordell-Weil theorem (Thm. 2.4.3), $E(\mathbb{Q}) \cong T \oplus \mathbb{Z}^{R_E}$ where $T = E(\mathbb{Q})_{\text{torsion}}$ is a finite abelian group. Therefore,

$$E(\mathbb{Q})/2E(\mathbb{Q}) \cong T/2T \oplus (\mathbb{Z}/2\mathbb{Z})^{R_E}.$$

In our restricted case, we have assumed all along that $E(\mathbb{Q})$ contains 4 points of 2-torsion, namely $\mathcal{O}$ and $(e_i, 0)$, for $i = 1, 2, 3$. And, by Exercise 2.12.6, $E(\mathbb{Q})$ cannot have more points of order 2. Thus, $T/2T \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$ (see Exercise 2.12.20).

Hence, the size of $E(\mathbb{Q})/2E(\mathbb{Q})$ is exactly 2^{R_E+2}, under our assumptions. Recall that we defined $\nu(N)$ to be the number of distinct prime divisors of an integer N. We prove our first bound:

Proposition 2.10.1. *Let $E : y^2 = (x - e_1)(x - e_2)(x - e_3)$ be an elliptic curve, with $e_i \in \mathbb{Z}$. Then the rank of $E(\mathbb{Q})$ is $R_E \leq 2\nu(\Delta_E)$.*

Proof. If the quantity Δ_E has $\nu = \nu(\Delta_E)$ distinct (positive) prime divisors, then we claim that the set

$$\Gamma' = \{n \in \mathbb{Z} : 0 \neq n \text{ is square-free and if } p \mid n, \text{ then } p \mid \Delta\}/(\mathbb{Z}^\times)^2$$

has precisely $2^{\nu(\Delta_E)+1}$ elements. Indeed, if $\Delta_E = p_1^{s_1} \cdots p_\nu^{s_\nu}$, then

$$\Gamma' = \{(-1)^{t_0} p_1^{t_1} \cdots p_\nu^{t_\nu} : t_i = 0 \text{ or } 1 \text{ for } i = 0, \ldots, \nu\}.$$

Thus, Γ' has as many elements as $\{(t_0, \ldots, t_\nu) : t_i = 0 \text{ or } 1\}$, which clearly has $2^{\nu+1}$ elements. Moreover, the set Γ_Δ, as defined in Corollary 2.9.6, has as many elements as $\Gamma' \times \Gamma'$, i.e., $2^{2\nu+2}$ elements. Since $E(\mathbb{Q})/2E(\mathbb{Q})$ injects into Γ_Δ, we conclude that it also has at most $2^{2\nu+2}$ elements. Since the size of $E(\mathbb{Q})/2E(\mathbb{Q})$ is 2^{R_E+2}, we conclude that $R_E + 2 \leq 2\nu + 2$ and $R_E \leq 2\nu$, as claimed. ∎

Example 2.10.2. Let

$$E : y^2 = x^3 - 1156x = x(x - 34)(x + 34).$$

The discriminant of $E/\mathbb{Q}$ is $\Delta_E = 98867482624 = 2^{12} \cdot 17^6$. Hence, $\nu(\Delta_E) = 2$ and the rank of E is at most 4. (The rank is in fact 2; see Example 2.10.4 below.) ∎

The bound $R_E \leq 2\nu(\Delta_E)$ is, in general, not very sharp (Theorem 2.7.4 is an improvement). However, the method we followed to come up with the bound yields a strategy to find generators for $E(\mathbb{Q})/2E(\mathbb{Q})$ as follows. Recall that $E(\mathbb{Q})/2E(\mathbb{Q})$ embeds into Γ_Δ via the map δ, so we want to identify which elements of Γ_Δ may belong to the image of δ. Suppose $(\delta_1, \delta_2, \delta_3) \in \Gamma_\Delta$ belongs to the image of δ and it is not the image of a torsion point. Then there exists $P = (x_0, y_0) \in E(\mathbb{Q})$ such that:

$$\begin{cases} y_0^2 = (x_0 - e_1)(x_0 - e_2)(x_0 - e_3), \\ x_0 - e_1 = \delta_1 u^2, \\ x_0 - e_2 = \delta_2 v^2, \\ x_0 - e_3 = \delta_3 w^2 \end{cases}$$

for some rational numbers u, v, w. We may substitute the last equation in the previous two, and obtain

$$\begin{cases} e_3 - e_1 = \delta_1 u^2 - \delta_3 w^2, \\ e_3 - e_2 = \delta_2 v^2 - \delta_3 w^2. \end{cases}$$

Recall that the elements $(\delta_1, \delta_2, \delta_3)$ that are in the image of δ satisfy $\delta_1 \cdot \delta_2 \cdot \delta_3 = 1$ modulo squares. Thus, $\delta_3 = \delta_1 \cdot \delta_2 \cdot \lambda^2$ and if we do a change of variables $(u, v, w) \mapsto (X, Y, \frac{Z}{\lambda})$, we obtain a system

$$C(\delta_1, \delta_2) : \begin{cases} e_3 - e_1 = \delta_1 X^2 - \delta_1 \delta_2 Z^2, \\ e_3 - e_2 = \delta_2 Y^2 - \delta_1 \delta_2 Z^2, \end{cases}$$

or, equivalently, one can subtract both equations to get

$$C(\delta_1,\delta_2): \begin{cases} e_1 - e_2 = \delta_2 Y^2 - \delta_1 X^2, \\ e_3 - e_2 = \delta_2 Y^2 - \delta_1\delta_2 Z^2. \end{cases}$$

The space $C(\delta_1,\delta_2)$ is the intersection of two conics, and it may have rational points or not. If $(\delta_1,\delta_2,\delta_3)$ is in the image of δ, however, then the space $C(\delta_1,\delta_2)$ must have a rational point; i.e., there are $X,Y,Z \in \mathbb{Q}$ that satisfy the equations of $C(\delta_1,\delta_2)$. Moreover, if $X_0,Y_0,Z_0 \in \mathbb{Q}$ are the coordinates of a point in $C(\delta_1,\delta_2)$, then

$$P = (e_1 + \delta_1 X_0^2,\ \delta_1\delta_2 X_0 Y_0 Z_0) \tag{2.12}$$

is a rational point on $E(\mathbb{Q})$ such that $\delta(P) = (\delta_1,\delta_2,\delta_3)$. The spaces $C(\delta_1,\delta_2)$ are called *homogeneous spaces* and are extremely helpful when we try to calculate the Mordell-Weil group of an elliptic curve. We record our findings in the form of a proposition, for later use:

Proposition 2.10.3. *Let $E/\mathbb{Q}$ be an elliptic curve with Weierstrass equation $y^2 = (x-e_1)(x-e_2)(x-e_3)$, with $e_i \in \mathbb{Z}$ and $e_1+e_2+e_3 = 0$. Let $\delta : E(\mathbb{Q})/2E(\mathbb{Q}) \hookrightarrow \Gamma_\Delta$ be the injection given by Corollary 2.9.7, and let $\delta(E) := \delta(E(\mathbb{Q})/2E(\mathbb{Q}))$ be the image of δ in Γ_Δ. Then:*

(1) *If $(\delta_1,\delta_2,\delta_3) \in \delta(E)$, then the space $C(\delta_1,\delta_2)$ has a point (X_0,Y_0,Z_0) with rational coordinates, $X_0,Y_0,Z_0 \in \mathbb{Q}$.*

(2) *Conversely, if $C(\delta_1,\delta_2)$ has a rational point (X_0,Y_0,Z_0), then $E(\mathbb{Q})$ has a rational point*

$$P = (e_1 + \delta_1 X_0^2,\ \delta_1\delta_2 X_0 Y_0 Z_0).$$

(3) *Since δ is a homomorphism and $\delta(E)$ is the image of δ, it follows that $\delta(E)$ is a subgroup of Γ_Δ. In particular:*
- *If $(\delta_1,\delta_2,\delta_3)$ and $(\delta_1',\delta_2',\delta_3')$ are elements of the image, then their product $(\delta_1\cdot\delta_1',\ \delta_2\cdot\delta_2',\ \delta_3\cdot\delta_3')$ is also in the image;*
- *If $(\delta_1,\delta_2,\delta_3) \in \delta(E)$ but $(\delta_1',\delta_2',\delta_3') \in \Gamma_\Delta$* **is not** *in the image, then their product $(\delta_1\cdot\delta_1',\ \delta_2\cdot\delta_2',\ \delta_3\cdot\delta_3')$* **is not** *in the image $\delta(E)$;*
- *If $C(\delta_1,\delta_2)$ and $C(\delta_1',\delta_2')$ have rational points, then $C(\delta_1\cdot\delta_1',\ \delta_2\cdot\delta_2')$ also has a rational point;*

- *If $C(\delta_1, \delta_2)$ has a rational point but $C(\delta_1', \delta_2')$* **does not have** *a rational point, then $C(\delta_1 \cdot \delta_1',\ \delta_2 \cdot \delta_2')$* **does not have** *a rational point.*

Example 2.10.4. Let $E : y^2 = x^3 - 1156x = x(x-34)(x+34)$. The only divisors of Δ_E are 2 and 17. Thus, $\Gamma' = \{\pm 1, \pm 2, \pm 17, \pm 34\}$. Let us choose $e_1 = 0$, $e_2 = -34$ and $e_3 = 34$. Therefore, the homogeneous spaces for this curve are all of the form

$$C(\delta_1, \delta_2) : \begin{cases} \delta_2 Y^2 - \delta_1 X^2 = 34, \\ \delta_2 Y^2 - \delta_1 \delta_2 Z^2 = 68 \end{cases}$$

with $\delta_1, \delta_2 \in \Gamma'$. We analyze these spaces, case by case. There are 64 pairs (δ_1, δ_2) to take care of:

(1) $((\delta_1, \delta_2, \delta_3) = (1,1,1))$. The point at infinity (i.e., the origin) is sent to $(1,1,1)$ via δ, i.e., $\delta(\mathcal{O}) = (1,1,1)$.

(2) ($\delta_1 < 0$ and $\delta_2 < 0$). The equation $\delta_2 Y^2 - \delta_1 \delta_2 Z^2 = 68$ cannot have solutions (in $\mathbb{Q}$ or $\mathbb{R}$) because the left-hand side is always negative for any $X, Z \in \mathbb{Q}$.

(3) ($\delta_1 > 0$ and $\delta_2 < 0$). The equation $\delta_2 Y^2 - \delta_1 X^2 = 34$ cannot have solutions (in $\mathbb{Q}$ or $\mathbb{R}$), because the left-hand side is always negative.

(4) ($\delta_1 = -1$, $\delta_2 = 34$). The space $C(-1, 34)$ has a rational point $(X, Y, Z) = (0, 1, 1)$, which maps to $T_1 = (0, 0)$ on $E(\mathbb{Q})$ via Eq. (2.12).

(5) ($\delta_1 = -34$, $\delta_2 = 2$). The space $C(-34, 2)$ has the rational point $(X, Y, Z) = (1, 0, 1)$, which maps to $T_2 = (-34, 0)$ on $E(\mathbb{Q})$ via Eq. (2.12).

(6) ($\delta_1 = 34$, $\delta_2 = 17$). If $\delta(T_1) = \delta((0,0))$ equals $(-1, 34, -34)$, and $\delta(T_2) = (-34, 2, -17)$, then

$$\delta(T_1 + T_2) = \delta(T_1) \cdot \delta(T_2) = (-1, 34, -34) \cdot (-34, 2, -17) = (34, 17, 2).$$

Thus, the space $C(34, 17)$ must have a point that maps back to $T_1 + T_2 = (34, 0)$. Indeed, $C(34, 17)$ has a point $(X, Y, Z) = (1, 2, 0)$ that maps to $(34, 0)$ via Eq. (2.12).

(7) ($\delta_1 = -1$, $\delta_2 = 2$). The space $C(-1,2)$ has a rational point $(X,Y,Z) = (4,3,5)$, which maps to $P = (-16,-120)$ on $E(\mathbb{Q})$ via Eq. (2.12). P is a point of infinite order.

(8) ($(\delta_1,\delta_2) = (1,17)$, $(34,1)$, or $(-34,34)$). These are the pairs that correspond to $(-1,2)\cdot\gamma$, with $\gamma = (-1,34)$, $(-34,2)$ or $(34,17)$. Therefore, the corresponding spaces $C(\delta_1,\delta_2)$ must have rational points that map to $P+T_1$, $P+T_2$ and $P+T_1+T_2$, respectively.

(9) ($\delta_1 = -2$, $\delta_2 = 2$). The space $C(-2,2)$ has a rational point $(X,Y,Z) = (1,4,3)$, which maps to $Q = (-2,-48)$ on $E(\mathbb{Q})$ via Eq. (2.12). Q is a point of infinite order.

(10) ($(\delta_1,\delta_2) = (2,17)$, $(17,1)$, or $(-17,34)$). These are the pairs that correspond to $(-2,2)\cdot\gamma$, with $\gamma = (-1,34)$, $(-34,2)$ or $(34,17)$. Therefore, the corresponding spaces $C(\delta_1,\delta_2)$ must have rational points that map to $Q+T_1$, $Q+T_2$ and $Q+T_1+T_2$, respectively.

(11) ($(\delta_1,\delta_2) = (2,1)$, and $(-2,34)$, $(-17,2)$, or $(17,17)$). Since $(-1,2)$ and $(-2,2)$ correspond to P and Q, respectively, then $(-1,2)\cdot(-2,2) = (2,1)$ corresponds to $P+Q$. The other pairs correspond to $(-2,2)\cdot\gamma$, with $\gamma = (-1,34)$, $(-34,2)$ or $(34,17)$. Therefore, the corresponding spaces $C(\delta_1,\delta_2)$ must have rational points that map to $P+Q+T_1$, $P+Q+T_2$ and $P+Q+T_1+T_2$, respectively.

(12) ($\delta_1 = 1$, $\delta_2 = 2$). The space $C(1,2)$ does not have rational points (see Exercise 2.12.21). In fact, it does not have solutions in $\mathbb{Q}_2$, the field of 2-adic numbers.

(13) ($(\delta_1,\delta_2) = (2,2)$, $(17,2)$, $(34,2)$, $(-1,1)$, $(-2,1)$, $(-17,1)$, $(-34,1)$, $(-1,17)$, $(-2,17)$, $(-17,17)$, $(-34,17)$, $(1,34)$, $(2,34)$, $(17,34)$, $(34,34)$). The corresponding spaces $C(\delta_1,\delta_2)$ do not have rational points. For instance, suppose $C(2,2)$ had a point. Then $(2,2,1)$ would be in the image of δ. Since $(2,1,2)$ *is* in the image of δ (we already saw above that $C(2,1)$ has a point), then $(2,1,2)\cdot(2,2,1) = (1,2,2)$ would also be in the image of δ, but we just saw (in the previous item) that $(1,2,2)$ is *not* in the image of δ. Therefore,

we have reached a contradiction and $C(2,2)$ cannot have a rational point. One can rule out all the other (δ_1, δ_2) in the list similarly.

We have analyzed all 64 possible pairs (δ_1, δ_2) and have found that the image of $E(\mathbb{Q})/2E(\mathbb{Q})$ via δ has order 2^4. Therefore, $2^{R_E+2} = 2^4$ and $R_E = 2$. The rank of the curve is exactly 2 and T_1, T_2, P and Q (as found above) are generators of $E(\mathbb{Q})/2E(\mathbb{Q})$. (In fact, they are generators of $E(\mathbb{Q})$ as well.) ■

Example 2.10.5. Let $E : y^2 = x^3 - 6724x = x(x-82)(x+82)$. Let $e_1 = 0$, $e_2 = -82$ and $e_3 = 82$. The only divisors of Δ_E are 2 and 41, hence $\Gamma' = \{\pm 1, \pm 2, \pm 41, \pm 82\}$. Let us analyze the homogeneous spaces

$$C(\delta_1, \delta_2) : \begin{cases} \delta_2 Y^2 - \delta_1 X^2 = 82, \\ \delta_2 Y^2 - \delta_1 \delta_2 Z^2 = 164 \end{cases}$$

as we did in the previous example. Once again, there are 64 pairs to check:

(1) $((\delta_1, \delta_2, \delta_3) = (1,1,1))$. The point at infinity (i.e., the origin) is sent to $(1,1,1)$ via δ, i.e., $\delta(\mathcal{O}) = (1,1,1)$.

(2) ($\delta_1 < 0$ and $\delta_2 < 0$). The equation $\delta_2 Y^2 - \delta_1 \delta_2 Z^2 = 164$ cannot have rational solutions because the left-hand side is always negative for any $X, Z \in \mathbb{Q}$.

(3) ($\delta_1 > 0$ and $\delta_2 < 0$). The equation $\delta_2 Y^2 - \delta_1 X^2 = 82$ cannot have rational solutions, because the left-hand side is always negative.

(4) $((\delta_1, \delta_2) = (-1, 82), (-82, 2), (82, 41))$. The corresponding spaces have (trivial) rational points that map, respectively, to $T_1 = (0,0)$, $T_2 = (-82, 0)$ and $T_3 = T_1 + T_2 = (82, 0)$ via Eq. (2.12).

(5) $((\delta_1, \delta_2) = (1, 2))$. The space $C(1,2)$ does not have rational points (same reason as for Exercise 2.12.21). In fact, it does not have any solutions over $\mathbb{Q}_2$.

(6) $((\delta_1, \delta_2) = (-1, 41), (-82, 1), (82, 82))$. The corresponding spaces cannot have rational points, because these elements of Γ_Δ are the product of $(1,2,2)$, with no points,

times $(-1, 82, -82)$, $(-82, 2, -41)$, $(82, 41, 2)$, which do have points by a previous item in this list.

How about all the other possible pairs (δ_1, δ_2)? Consider $(-1, 2, -2)$ and its homogeneous space:

$$C(-1,2): \begin{cases} 2Y^2 + X^2 = 82, \\ 2Y^2 + 2Z^2 = 164. \end{cases}$$

Let us show that there are solutions to $C(-1,2)$ over $\mathbb{R}$, $\mathbb{Q}_2$ and $\mathbb{Q}_{41}$:

- (Over $\mathbb{R}$). The point $(0, \sqrt{41}, \sqrt{41})$ is a point on $C(-1,2)$ defined over $\mathbb{R}$.
- (Over $\mathbb{Q}_{41}$). Let $Y_0 = 1$ and put $f(X) = X^2 - 80$, $g(Z) = Z^2 - 81$. By Hensel's Lemma (see Appendix D.1 and Corollary D.1.2), it suffices to show that there are $\alpha_0, \beta_0 \in \mathbb{F}_{41}$ such that

$$f(\alpha_0) = g(\beta_0) \equiv 0 \bmod 41 \quad \text{and} \quad f'(\alpha_0),\ g'(\beta_0) \not\equiv 0 \bmod 41.$$

  The reader can check that the congruences $\alpha_0 \equiv 11 \bmod 41$ and $\beta_0 \equiv 9 \bmod 41$ work. Thus, there are $\alpha, \beta \in \mathbb{Q}_{41}$ such that $f(\alpha) = 0 = g(\beta)$. Hence, $(X_0, Y_0, Z_0) = (\alpha, 1, \beta)$ is a point on $C(-1,2)$ defined over $\mathbb{Q}_{41}$, as desired.
- (Over $\mathbb{Q}_2$). Let $X_0 = 0$ and put $f(Y) = Y^2 - 41$. Let $\alpha_0 = 1$. Then $f(\alpha_0) = -40$, $f'(\alpha_0) = 82$ and

$$3 = \nu_2(-40) > \nu_2(82^2) = \nu_2(2^2 \cdot 41^2) = 2.$$

  Thus, by Hensel's Lemma (Theorem D.1.1; see also Example D.1.4), there is $\alpha \in \mathbb{Q}_2$ such that $f(\alpha) = 0$, or $\alpha^2 = 41$. Hence, the point $(X_0, Y_0, Z_0) = (0, \alpha, \alpha)$ is a point on $C(-1,2)$ defined over $\mathbb{Q}_2$, as desired.

One can also show that, in fact, $C(-1,2)$ has a point over $\mathbb{Q}_p$ *for all* $p \geq 2$. Therefore, we cannot deduce any contradictions working locally about whether $C(-1,2)$ has a point over $\mathbb{Q}$. A computer search does not yield any $\mathbb{Q}$-points on $C(-1,2)$. Therefore, our method breaks at this point, and we cannot determine whether there is a point on $E(\mathbb{Q})$ that comes from $C(-1,2)$.

It turns out that $C(-1,2)$ *does not* have rational points (but this is difficult to show). This type of space, a space that has solutions

everywhere locally ($\mathbb{Q}_p$, $\mathbb{R}$) but not globally ($\mathbb{Q}$) is the main obstacle for the descent method to fully work. ■

2.11. Selmer and Sha

In Example 2.10.5, we found a type of homogeneous space that made our approach to finding generators of $E(\mathbb{Q})/2E(\mathbb{Q})$ break down. In this section, we study everywhere locally solvable spaces in detail.

Let $E : y^2 = (x - e_1)(x - e_2)(x - e_3)$ be an elliptic curve with $e_i \in \mathbb{Z}$ and $e_1 + e_2 + e_3 = 0$. Let Γ' be defined as in Corollary 2.9.7, i.e.:

$$\Gamma' = \{n \in \mathbb{Z} : 0 \neq n \text{ is square-free and if } p \mid n, \text{ then } p \mid \Delta\}/(\mathbb{Z}^\times)^2$$

where $\Delta = (e_1 - e_2)(e_2 - e_3)(e_1 - e_3)$. We define $\mathcal{H}$ as the following set of homogeneous spaces:

$$\mathcal{H} := \{C(\delta_1, \delta_2) : \delta_1, \delta_2 \in \Gamma'\}.$$

Some homogeneous spaces in $\mathcal{H}$ have rational points that correspond to rational points on $E(\mathbb{Q})$; see Prop. 2.10.3. Other homogeneous spaces do not have points (e.g. $C(1, 2)$ in Example 2.10.4, or $C(-1, 2)$ in Example 2.10.5). For each elliptic curve, we define two different sets of homogeneous spaces, the Selmer group and the Shafarevich-Tate group, as follows. The *Selmer group* is

$$\mathrm{Sel}_2(E/\mathbb{Q}) := \{C(\delta_1, \delta_2) \text{ with points over } \mathbb{R} \text{ and } \mathbb{Q}_p \text{ for all primes } p\}.$$

In other words, the Selmer group is the set of all homogeneous spaces that are solvable everywhere *locally*, i.e., over $\mathbb{R}$ and over all fields of p-adic numbers. The group operation on $\mathrm{Sel}_2(E/\mathbb{Q})$ is defined by

$$C(\delta_1, \delta_2) \cdot C(\gamma_1, \gamma_2) = C(\delta_1\gamma_1, \delta_2\gamma_2).$$

Notice that, due to Prop. 2.10.3, $E(\mathbb{Q})/2E(\mathbb{Q})$ injects into $\mathcal{H}$ via δ and the homogeneous spaces in the image of δ, i.e. $\delta(E) \subseteq \mathcal{H}$, have rational points. Since $\mathbb{Q} \subseteq \mathbb{Q}_p$ for all primes $p \geq 2$, the spaces in the image of δ belong to $\mathrm{Sel}_2(E/\mathbb{Q})$. Hence, $\mathrm{Sel}_2(E/\mathbb{Q})$ has a subgroup formed by those homogeneous spaces in $\mathrm{Sel}_2(E/\mathbb{Q})$ that have rational points as well (i.e., over $\mathbb{Q}$), and this subgroup is isomorphic to $E(\mathbb{Q})/2E(\mathbb{Q})$:

$$E(\mathbb{Q})/2E(\mathbb{Q}) = \{C(\delta_1, \delta_2) \text{ with points defined over } \mathbb{Q}\}.$$

Finally, the *Shafarevich-Tate group* is the *quotient* of the Selmer group by its subgroup $E(\mathbb{Q})/2E(\mathbb{Q})$. Thus, each element of the Shafarevich-Tate group is represented by $C(1,1)$ or by a homogeneous space that is solvable everywhere locally but *does not* have a rational point:

$$\begin{aligned} Ш_2(E/\mathbb{Q}) &= \{C(1,1)\} \\ &\quad \cup \{C(\delta_1,\delta_2) \in \mathrm{Sel}_2(E/\mathbb{Q}) \text{ without points over } \mathbb{Q}\}. \end{aligned}$$

These three groups, Selmer, Ш (or "Sha") and $E/2E$, fit in a *short exact sequence*

$$0 \longrightarrow E(\mathbb{Q})/2E(\mathbb{Q}) \longrightarrow \mathrm{Sel}_2(E/\mathbb{Q}) \longrightarrow Ш_2(E/\mathbb{Q}) \longrightarrow 0.$$

In other words, the map $\psi : E(\mathbb{Q})/2E(\mathbb{Q}) \to \mathrm{Sel}_2(E/\mathbb{Q})$ is injective, the map $\phi : \mathrm{Sel}_2(E/\mathbb{Q}) \to Ш_2(E/\mathbb{Q})$ is surjective, and the kernel of ϕ is the image of ψ.

Remark 2.11.1. Here for simplicity we have defined what number theorists would usually refer to as the 2-part of the Selmer group ($\mathrm{Sel}_2(E/\mathbb{Q})$ above) and the 2-torsion of Ш (the group $Ш_2$ as defined above). For the definition of the full Selmer and Ш groups, see [**Sil86**], Ch. X, §4.

Example 2.11.2. Let $E : y^2 = x^3 - 1156x$, as in Example 2.10.4. The full group of homogeneous spaces $\mathcal{H}$ has 64 elements:

$$\mathcal{H} = \{C(\delta_1,\delta_2) : \delta_i = \pm 1, \pm 2, \pm 17, \pm 34\}.$$

The spaces in $\mathcal{H}$ with $\delta_2 < 0$ do not have points over $\mathbb{R}$, so they do not belong to $\mathrm{Sel}_2(E/\mathbb{Q})$. Moreover, we showed that the spaces $(\delta_1,\delta_2) = (2,2)$, $(17,2)$, $(34,2)$, $(-1,1)$, $(-2,1)$, $(-17,1)$, $(-34,1)$, $(-1,17)$, $(-2,17)$, $(-17,17)$, $(-34,17)$, $(1,34)$, $(2,34)$, $(17,34)$, and $(34,34)$ do not have points over $\mathbb{Q}_2$. Therefore, they do not belong to $\mathrm{Sel}_2(E/\mathbb{Q})$ either. All other spaces have rational points; therefore, they are everywhere locally solvable, so they all belong to $\mathrm{Sel}_2(E/\mathbb{Q})$. Hence,

$$\begin{aligned} \mathrm{Sel}_2(E/\mathbb{Q}) &= \{C(\delta_1,\delta_2) : (\delta_1,\delta_2) = \\ &\qquad (1,1),(-1,34),(-34,2),(34,17), \\ &\qquad (1,17),(34,1),(-34,34),(-2,2), \\ &\qquad (17,1),(-17,34),(2,1),(-2,34), \\ &\qquad (-17,2),(17,17),(-1,2),(2,17)\}. \end{aligned}$$

Notice that, indeed, the elements of $\mathrm{Sel}_2(E/\mathbb{Q})$ listed above form a subgroup of $\Gamma' \times \Gamma' \subset (\mathbb{Q}^\times/(\mathbb{Q}^\times)^2)^2$. Since all the elements of $\mathrm{Sel}_2(E/\mathbb{Q})$ have rational points, we conclude that $\mathrm{Sel}_2(E/\mathbb{Q})$ equals $E(\mathbb{Q})/2E(\mathbb{Q})$ and

$$Ш_2(E/\mathbb{Q}) = \mathrm{Sel}_2(E/\mathbb{Q})/(E(\mathbb{Q})/2E(\mathbb{Q})) = \{C(1,1)\},$$

i.e., $Ш_2$ is the trivial subgroup in this case. ■

Example 2.11.3. Let $E : y^2 = x^3 - 6724x$, as in Example 2.10.5. The full group of homogeneous spaces $\mathcal{H}$ has 64 elements:

$$\mathcal{H} = \{C(\delta_1, \delta_2) : \delta_i = \pm 1, \pm 2, \pm 41, \pm 82\}.$$

The spaces in $\mathcal{H}$ with $\delta_2 < 0$ do not have points over $\mathbb{R}$, so they do not belong to $\mathrm{Sel}_2(E/\mathbb{Q})$. Moreover, the spaces $(\delta_1, \delta_2) = (2,2)$, $(41,2)$, $(82,2)$, $(-1,1)$, $(-2,1)$, $(-41,1)$, $(-82,1)$, $(-1,41)$, $(-2,41)$, $(-41,41)$, $(-82,41)$, $(1,82)$, $(2,82)$, $(41,82)$, and $(82,82)$ do not have points over $\mathbb{Q}_2$. Therefore, they do not belong to $\mathrm{Sel}_2(E/\mathbb{Q})$ either. It turns out that the rest of the spaces (such as $C(-1,2)$) are everywhere locally solvable (we showed this for $C(-1,2)$). Therefore they all belong to $\mathrm{Sel}_2(E/\mathbb{Q})$. Hence,

$$\begin{aligned}\mathrm{Sel}_2(E/\mathbb{Q}) \quad = \quad \{C(\delta_1,\delta_2) : (\delta_1,\delta_2) = \\ (1,1), (-1,82), (-82,2), (82,41), \\ (1,41), (82,1), (-82,82), (-2,2), \\ (41,1), (-41,82), (2,1), (-2,82), \\ (-41,2), (41,41), (-1,2), (2,41)\}.\end{aligned}$$

The spaces $(1,1)$, $(-1,82)$, $(-82,2)$ and $(82,41)$ have rational points that correspond to (torsion) points on $E(\mathbb{Q})$. However, *none* of the other spaces have rational solutions! Thus, the rest are representative of non-trivial elements of Sha, and we conclude that

$$E(\mathbb{Q})/2E(\mathbb{Q}) = \{C(1,1), C(-1,82), C(-82,2), C(82,41)\}$$

and $Ш_2(E/\mathbb{Q}) = \{C(\delta_1,\delta_2) : (\delta_1,\delta_2) = (1,1), (-1,2), (-2,2), (2,1)\}$.

Notice that the elements of $Ш_2$ listed above are representatives of all the classes in the quotient of $\mathrm{Sel}_2(E/\mathbb{Q})$ by $E(\mathbb{Q})/2E(\mathbb{Q})$. For instance, $(-1,2)\cdot(1,41) = (-1,82) \in E(\mathbb{Q})/2E(\mathbb{Q})$. Thus, $(-1,2)\cdot(1,41)$ is trivial in $Ш_2$. ■

2.12. Exercises

Exercise 2.12.1. Let $f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n$, with $a_i \in \mathbb{Z}$. Prove that if $x = \frac{p}{q} \in \mathbb{Q}$, with $\gcd(p, q) = 1$, is a solution of $f(x) = 0$, then a_n is divisible by p and a_0 is divisible by q.

Exercise 2.12.2. Let C be the conic defined by $x^2 - 2y^2 = 1$.

(1) Find all the rational points on C. (Hint: the point $O = (1, 0)$ belongs to C. Let $L(t)$ be the line that goes through O and has slope t. Since C is a quadratic and $L(t) \cap C$ contains at least one rational point, there must be a second point of intersection Q. Find the coordinates of Q in terms of t.)

(2) Let $\alpha = 1 + \sqrt{2}$. Calculate $\alpha^2 = a + b\sqrt{2}$ and $\alpha^4 = c + d\sqrt{2}$ and verify that (a, b) and (c, d) are integral points on C : $x^2 - 2y^2 = 1$. (Note: in fact, if $\alpha^{2n} = e + f\sqrt{2}$, then $(e, f) \in C$ and the coefficients of α^{2n+1} are a solution of $x^2 - 2y^2 = -1$.)

(3) (This problem is only for those who already know about continued fractions.) Find the continued fraction of $\sqrt{2}$ and find the first 6 convergents. Use the convergents to find three distinct (positive) integral solutions of $x^2 - 2y^2 = 1$, other than $(1, 0)$. (Note: the reader should remind herself or himself how to find the continued fraction and convergents *by hand*, then check his or her answer using Sage; see Appendix A.4.)

Exercise 2.12.3. Let $C/\mathbb{Q}$ be an affine curve.

(1) Suppose that $C/\mathbb{Q}$ is given by an equation of the form

$$C : xy^2 + ax^2 + bxy + cy^2 + dx + ey + f = 0. \tag{2.13}$$

Find an invertible change of variables that takes the equation of C onto one of the form $xy^2+gx^2+hxy+jx+ky+l = 0$. (Hint: consider a change of variables $X = x + \lambda$, $Y = y$).

(2) Suppose that $C'/\mathbb{Q}$ is given by an equation of the form

$$C' : xy^2 + ax^2 + bxy + cx + dy + e = 0. \tag{2.14}$$

Find an invertible change of variables that takes the equation of C' onto one of the form $y^2 + \alpha xy + \beta y = x^3 + \gamma x^2 +$

$\delta x + \eta$. (Hint: multiply (2.14) by x and consider the change of variables $X = x$ and $Y = xy$. Make sure that, at the end, the coefficients of y^2 and x^3 equal 1.)

(3) Suppose that $C''/\mathbb{Q}$ is a curve given by an equation of the form

$$C'' : y^2 + axy + by = x^3 + cx^2 + dx + e. \tag{2.15}$$

Find an invertible change of variables that takes the equation of C'' onto one of the form $y^2 = x^3 + Ax + B$. (Hint: do it in two steps. First eliminate the xy and y terms. Then eliminate the x^2 term.)

(4) Let $E/\mathbb{Q} : y^2 + 43xy - 210y = x^3 - 210x^2$. Find an invertible change of variables that takes the equation of E to one of the form $y^2 = x^3 + Ax + B$.

Exercise 2.12.4. Let C and E be curves defined, respectively, by $C : V^2 = U^4 + 1$ and $E : y^2 = x^3 - 4x$. Let ψ be the map defined by

$$\psi(U, V) = \left(\frac{2(V+1)}{U^2}, \frac{4(V+1)}{U^3} \right).$$

(1) Show that if $U \neq 0$ and $(U, V) \in C(\mathbb{Q})$, then $\psi(U, V) \in E(\mathbb{Q})$.

(2) Find an inverse function for ψ; i.e., find $\varphi : E \to C$ such that $\varphi(\psi(U, V)) = (U, V)$.

Next, we work in projective coordinates. Let $C : W^2V^2 = U^4 + W^4$ and $E : zy^2 = x^3 + z^3$.

(3) Write down the definition of ψ in projective coordinates; i.e., what is $\psi([U, V, W])$?

(4) Show that $\psi([0, 1, 1]) = [0, 1, 0] = \mathcal{O}$.

(5) Show that $\psi([0, -1, 1]) = [0, 0, 1]$. (*Hint: Show that*

$$\psi([U, V, W]) = [2U^2, 4UW, W(V - W)].)$$

Exercise 2.12.5. Use Sage to solve the following problems:

(1) Find $3Q$, where $E : y^2 = x^3 - 25x$ and $Q = (-4, 6)$. Use $3Q$ to find a new right triangle with rational sides and area equal to 5. (Hint: Examples 1.1.2 and 2.4.1.)

(2) Let $y^2 = x(x+5)(x+10)$ and $P = (-9, 6)$. Find nP for $n = 1, \ldots, 12$. Compare the x-coordinates of nP with the list given at the end of Example 1.1.1, and write down the next three numbers that belong in the list.

Exercise 2.12.6. Let $E/\mathbb{Q}$ be an elliptic curve given by a Weierstrass equation of the form $y^2 = f(x)$, where $f(x) \in \mathbb{Z}[x]$ is a monic cubic polynomial with distinct roots (over $\mathbb{C}$).

(1) Show that $P = (x, y) \in E$ is a torsion point of exact order 2 if and only if $y = 0$ and $f(x) = 0$.

(2) Let $E(\mathbb{Q})[2]$ be the subgroup of $E(\mathbb{Q})$ formed by those rational points $P \in E(\mathbb{Q})$ such that $2P = \mathcal{O}$. Show that the size of $E(\mathbb{Q})[2]$ may be 1, 2 or 4.

(3) Give examples of three elliptic curves defined over $\mathbb{Q}$ where the size of $E(\mathbb{Q})[2]$ is 1, 2 and 4, respectively.

Exercise 2.12.7. Let $E_t : y^2 + (1-t)xy - ty = x^3 - tx^2$ with $t \in \mathbb{Q}$ and $\Delta_t = t^5(t^2 - 11t - 1) \neq 0$. As we saw in Example 2.5.4 (or Appendix E), every curve E_t has a subgroup isomorphic to $\mathbb{Z}/5\mathbb{Z}$. Use Sage to find elliptic curves with torsion $\mathbb{Z}/5\mathbb{Z}$ and rank 0, 1 and 2. Also, try to find an elliptic curve E_t with rank r, as high as possible. (Note: the highest rank known — as of 6/1/2009 — for an elliptic curve with $\mathbb{Z}/5\mathbb{Z}$ torsion is 6, discovered by Dujella and Lecacheux in 2001; see [**Duj09**].)

Exercise 2.12.8. Let $p \geq 2$ be a prime and $E_p : y^2 = x^3 + p^2$. Show that there is no torsion point $P \in E_p(\mathbb{Q})$ with $y(P)$ equal to

$$y = \pm 1,\ \pm p^2,\ \pm 3p,\ \pm 3p^2,\ \text{or}\ \pm 3.$$

Prove that $Q = (0, p)$ is a torsion point of exact order 3. Conclude that $\{\mathcal{O}, Q, 2Q\}$ are the only torsion points on $E_p(\mathbb{Q})$. (Note: for $p = 3$, the point $(-2, 1) \in E_3(\mathbb{Q})$. Show that it is *not* a torsion point.)

Exercise 2.12.9. Prove Proposition 2.6.8, as follows:

(1) First show that if $f(x)$ is a polynomial, $f'(x)$ its derivative, and $f(\delta) = f'(\delta) = 0$, then $f(x)$ has a double root at δ.

(2) Show that if $y^2 = f(x)$ is singular, where $f(x) \in K[x]$ is a monic cubic polynomial, then the singularity must occur at $(\delta, 0)$, where δ is a root of $f(x)$.

(3) Show that $(\delta, 0)$ is singular if and only if δ is a double root of $f(x)$. Therefore $D = 0$ if and only if E is singular.

Exercise 2.12.10. Let $E/\mathbb{Q} : y^2 = x^3 + 3$. Find all the points of $\widetilde{E}(\mathbb{F}_7)$ and verify that N_7 satisfies Hasse's bound.

Exercise 2.12.11. Let $E/\mathbb{Q} : y^2 = x^3 + Ax + B$ and let $p \geq 3$ be a prime of bad reduction for $E/\mathbb{Q}$. Show that $E(\mathbb{F}_p)$ has a unique singular point.

Exercise 2.12.12. Prove parts (1) and (3) of Theorem 2.8.5. (Hint: use Definition 2.8.4 and Proposition 2.7.3.)

Exercise 2.12.13. Prove Corollary 2.8.6.

Exercise 2.12.14. Let $E : y^2 = x^3 - 10081x$. Use Sage (or PARI) to find a minimal set of generators for the subgroup that is spanned by all these points on E:

$$(0,0),\ (-100,90),\ \left(\frac{10081}{100}, \frac{90729}{1000}\right),\ (-17,408)$$

$$\left(\frac{907137}{6889}, -\frac{559000596}{571787}\right),\ \left(\frac{1681}{16}, \frac{20295}{64}\right), \left(\frac{833}{4}, \frac{21063}{8}\right)$$

$$\left(-\frac{161296}{1681}, \frac{19960380}{68921}\right), \left(-\frac{6790208}{168921}, -\frac{40498852616}{69426531}\right).$$

(Hint: use Theorem 2.7.4 to determine the rank of $E/\mathbb{Q}$.)

Exercise 2.12.15. Let E and δ be defined as in Theorem 2.9.3, and suppose $P = (x_0, y_0)$ is a point on E with $y_0 \neq 0$. Show:

- $\delta(P) \cdot \delta(\mathcal{O}) = \delta(P)$.
- $\delta((e_1, 0)) \cdot \delta((e_2, 0)) = \delta((e_1, 0) + (e_2, 0))$.
- $\delta(P) \cdot \delta((e_1, 0)) = \delta(P + (e_1, 0))$.

Exercise 2.12.16. Let $E : y^2 = x^3 + Ax + B$ be an elliptic curve with $A, B \in \mathbb{Q}$, and suppose $P = (x_0, y_0)$ is a point on E, with $y_0 \neq 0$.

(1) Prove that the x-coordinate of $2P$ is given by

$$x(2P) = \frac{x_0^4 - 2Ax_0^2 - 8Bx_0 + A^2}{4y_0^2}.$$

(2) Find a formula for $y(2P)$ in terms of x_0 and y_0.

Exercise 2.12.17. The curve $E/\mathbb{Q} : y^2 = x^3 - 157^2 x$ has a rational point Q with x-coordinate $x = x(Q)$ given by

$$x = \left(\frac{22440351770433696992455751309067486316094847204 1}{17824664537857719176051070357934327140032961660} \right)^2.$$

Show that there exists a point $P \in E(\mathbb{Q})$ such that $2P = Q$. Find the coordinates of P. (Hint: use PARI or Sage and Exercise 2.12.16.)

Exercise 2.12.18. Let $E : y^2 = (x-e_1)(x-e_2)(x-e_3)$ with $e_i \in \mathbb{Q}$, distinct, and such that $e_1 + e_2 + e_3 = 0$. Additionally, suppose that $e_1 - e_2 = n^2$ and $e_2 - e_3 = m^2$ are squares. This exercise shows that, under these assumptions, there is a point $P = (x_0, y_0)$ such that $2P = (e_1, 0)$, i.e., P is a point of exact order 4.

(1) Show that $e_1 = \frac{n^2+m^2}{3}$, $e_2 = \frac{m^2-2n^2}{3}$, $e_3 = \frac{n^2-2m^2}{3}$.

(2) Find A and B, in terms of n and m, such that $x^3 + Ax + B = (x-e_1)(x-e_2)(x-e_3)$. (Hint: Sage or PARI can be of great help here.)

(3) Let $p(x) = x^4 - 2Ax^2 - 8Bx + A^2 - 4(x^3 + Ax + B)e_1$. Show that $p(x_0) = 0$ if and only if $x(2P) = e_1$, and therefore $2P = (e_1, 0)$. (Hint: use Exercise 2.12.16.)

(4) Express all the coefficients of $p(x)$ in terms of n and m. (Hint: use Sage or PARI.)

(5) Factor $p(x)$ for $(n, m) = (3, 6),\ (3, 12),\ (9, 12), \ldots$.

(6) Guess that $p(x) = (x-a)^2(x-b)^2$ for some a and b. Express all the coefficients of $p(x)$ in terms of a and b.

(7) Finally, compare the coefficients of $p(x)$ in terms of a, b and n, m and find the roots of $p(x)$ in terms of n, m. (Hint: compare first the coefficient of x^3 and then the coefficient of x^2.)

(8) Write $P = (x_0, y_0)$ in terms of n and m.

Exercise 2.12.19. Let e_1, e_2, e_3 be three distinct integers. Show that $\Delta = (e_1 - e_2)(e_2 - e_3)(e_1 - e_3)$ is always even.

Exercise 2.12.20. In this exercise we study the structure of the quotient $G/2G$, where G is a finite abelian group.

(1) Let $p \geq 2$ be a prime and let $G = \mathbb{Z}/p^e\mathbb{Z}$, with $e \geq 1$. Prove that $G/2G$ is trivial if and only if $p > 2$.

(2) Prove that, if $G = \mathbb{Z}/2^e\mathbb{Z}$ and $e \geq 1$, then $G/2G \cong \mathbb{Z}/2\mathbb{Z}$.

(3) Finally, let G be an arbitrary finite abelian group. We define $G[2^\infty]$ to be the 2-primary component of G, i.e.,

$$G[2^\infty] = \{g \in G : 2^n \cdot g = 0 \text{ for some } n \geq 1\}.$$

In other words, $G[2^\infty]$ is the subgroup of G formed by those elements of G whose order is a power of 2. Prove that

$$G[2^\infty] \cong \mathbb{Z}/2^{e_1}\mathbb{Z} \oplus \mathbb{Z}/2^{e_2}\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/2^{e_r}\mathbb{Z}$$

for some $r \geq 0$ and $e_i \geq 1$ (here $r = 0$ means $G[2^\infty]$ is trivial). Also show that $G/2G \cong (\mathbb{Z}/2\mathbb{Z})^r$.

Exercise 2.12.21. Show that the space

$$C : \begin{cases} 2Y^2 - X^2 = 34, \\ Y^2 - Z^2 = 34 \end{cases}$$

does not have any rational solutions with $X, Y, Z \in \mathbb{Q}$. (Hint: modify the system so there are no powers of 2 in any of the denominators, then work modulo 8.)

Exercise 2.12.22. For the following elliptic curves, use the method of 2-descent (as in Proposition 2.10.3 and Example 2.10.4) to find the rank of $E/\mathbb{Q}$ and generators of $E(\mathbb{Q})/2E(\mathbb{Q})$. **Do not** use Sage:

(1) $E : y^2 = x^3 - 14931x + 220590$.

(2) $E : y^2 = x^3 - x^2 - 6x$.

(3) $E : y^2 = x^3 - 37636x$.

(4) $E : y^2 = x^3 - 962x^2 + 148417x$. (Hint: use Theorem 2.7.4 first to find a bound on the rank.)

Exercise 2.12.23. Find the rank and generators for the rational points on the elliptic curve $y^2 = x(x + 5)(x + 10)$.

Exercise 2.12.24. (Elliptic curves with non-trivial rank.) The goal here is a systematic way to find curves of rank at least $r \geq 0$ without using tables of elliptic curves:

(1) (Easy) Find 3 non-isomorphic elliptic curves over $\mathbb{Q}$ with rank ≥ 2. You must prove that the rank is at least 2. (To show linear independence, you may use PARI or Sage to calculate the height matrix).

(2) (Fair) Find 3 non-isomorphic elliptic curves over $\mathbb{Q}$ with rank ≥ 3.

(3) (Medium difficulty) Find 3 non-isomorphic elliptic curves over $\mathbb{Q}$ with rank ≥ 6. If so, then you can probably find 3 curves of rank ≥ 8 as well.

(4) (Significantly harder) Find 3 non-isomorphic elliptic curves over $\mathbb{Q}$ of rank ≥ 10.

(5) (You would be famous!) Find an elliptic curve over $\mathbb{Q}$ of rank ≥ 29.

Exercise 2.12.25. Let E be an elliptic curve and suppose that the images of the points $P_1, P_2, \ldots, P_n \in E(\mathbb{Q})$ in $E(\mathbb{Q})/2E(\mathbb{Q})$ generate the group $E(\mathbb{Q})/2E(\mathbb{Q})$. Let G be the subgroup of $E(\mathbb{Q})$ generated by $P_1, P_2, \ldots, P_n$.

(1) Prove that the index of G in $E(\mathbb{Q})$ is finite, i.e., the quotient group $E(\mathbb{Q})/G$ is finite.

(2) Show that, depending on the choice of generators $\{P_i\}$ of the quotient $E(\mathbb{Q})/2E(\mathbb{Q})$, the size of $E(\mathbb{Q})/G$ may be arbitrarily large.

Exercise 2.12.26. Fermat's last theorem shows that $x^3 + y^3 = z^3$ has no integer solutions with $xyz \neq 0$. Find the first $d \geq 1$ such that $x^3 + y^3 = dz^3$ has infinitely many non-trivial solutions, find a generator for the solutions and write down a few examples. (Hint: Example 2.2.3.)

Chapter 3

Modular curves

We saw in the introduction (Section 1.2) that a modular form is an object defined analytically. So far, we have only discussed algebraic aspects of elliptic curves. Before we go into the precise definitions of modular forms (Chapter 4), we need to consider elliptic curves over the complex numbers in order to motivate the definition of modular curves from the theory of elliptic curves, which in turn will motivate the definition of modular forms. In this chapter, we shall see that when we consider an elliptic curve $E/\mathbb{Q}$ as defined over $\mathbb{C}$ instead, then $E(\mathbb{C})$ is homeomorphic to a torus over $\mathbb{C}$. We remind the reader that Appendix B contains a concise introduction to complex analysis.

3.1. Elliptic curves over $\mathbb{C}$

3.1.1. Lattices.

Definition 3.1.1. A lattice L in the complex plane is an additive discrete subgroup of $\mathbb{C}$ such that $L \otimes \mathbb{R} = \mathbb{C}$.

Alternatively, a lattice can be defined by its generators. Let $w_1 = u_1 + v_1 i$ and $w_2 = u_2 + v_2 i$ be two non-zero complex numbers such that the vectors (u_1, v_1) and (u_2, v_2) are linearly independent in $\mathbb{R}^2$. Then, the set

$$L = \{mw_1 + nw_2 : m, n \in \mathbb{Z}\}$$

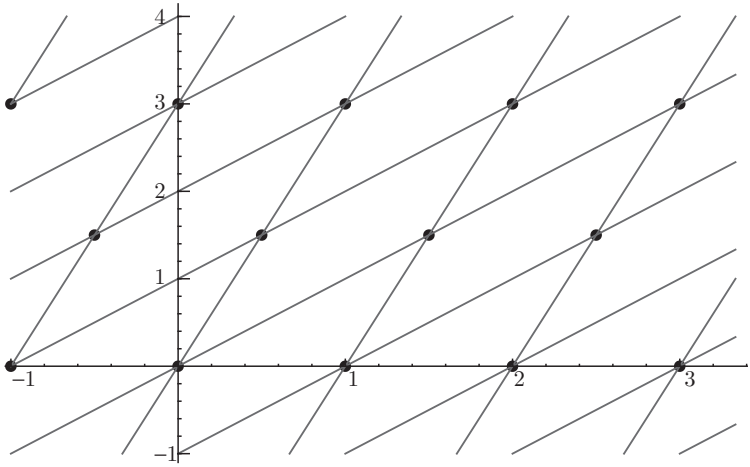


Figure 1. Points in the lattice $\langle \frac{1}{2} + \frac{3i}{2}, \frac{3}{2} + \frac{3i}{2} \rangle$.

is a lattice, and every lattice is given in this way. The lattice generated by $w_1, w_2 \in \mathbb{C}$ is denoted by $\langle w_1, w_2 \rangle$. We will insist on a *positive orientation* of our basis; i.e., we require w_1/w_2 to have positive imaginary part. In other words, w_1/w_2 belongs to the upper half complex plane $\mathbb{H}$, where

$$\mathbb{H} = \{a + bi \in \mathbb{C} : b > 0\}.$$

Example 3.1.2. The Gaussian integers $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$ form a lattice. One can take $w_1 = i$ and $w_2 = 1$ as generators (notice that w_1/w_2 has positive imaginary part). See Exercise 3.7.2. ■

Example 3.1.3. The set $\mathbb{Z}[\sqrt{2}] = \{a+b\sqrt{2} : a, b \in \mathbb{Z}\}$ *is not* a lattice, because when you replace $a, b \in \mathbb{Z}$ by $a, b \in \mathbb{R}$ we do not obtain all of $\mathbb{C}$ but only a 1-dimensional real space (in this case just $\mathbb{R}$). In other words, there are no two points w_1, w_2 in $\mathbb{Z}(\sqrt{2})$ whose coordinates are linearly independent in $\mathbb{R}^2$. ■

We shall be interested in quotients of $\mathbb{C}$ by a lattice L.

Definition 3.1.4. Let L be a lattice, and let $w_1, w_2 \in \mathbb{C}$ be generators of L. The group $\mathbb{C}/L$ is the quotient of $\mathbb{C}$, as an additive group, by

its subgroup L. In other words, we define $\mathbb{C}/L$ via an equivalence relation: we say that z_1 and z_2 are equivalent modulo L if there is $w \in L$ such that $z_1 - z_2 = w$. Then $\mathbb{C}/L$ is the set of equivalence classes of $\mathbb{C}$ modulo L.

If $L = \langle w_1, w_2 \rangle$, then the parallelogram

$$\mathcal{F} = \{\lambda w_1 + \mu w_2 : 0 \leq \lambda, \mu < 1\}$$

is called a *fundamental domain* for $\mathbb{C}/L$. Notice that there is a one-to-one correspondence between elements of $\mathcal{F}$ and classes in $\mathbb{C}/L$; i.e., the elements of $\mathcal{F}$ form a complete set of representatives for $\mathbb{C}/L$.

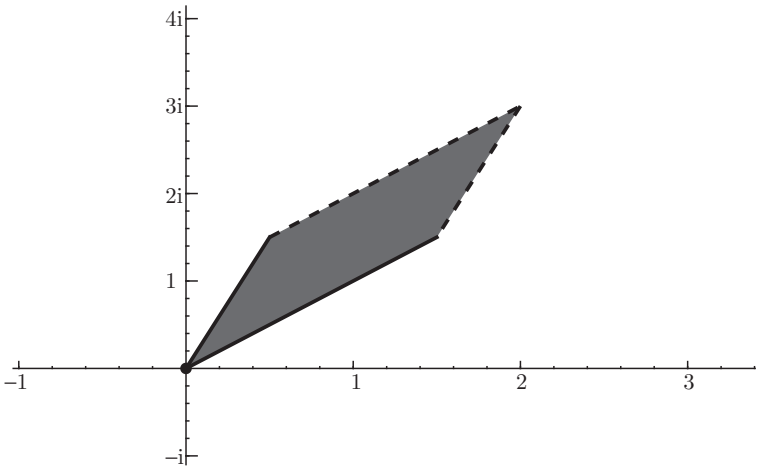


Figure 2. A fundamental domain for the lattice $\langle \frac{1}{2} + \frac{3i}{2}, \frac{3}{2} + \frac{3i}{2} \rangle$.

Notice also that if L is a lattice, then $\mathbb{C}/L$ is a (flat) torus because each side of the parallelogram $\mathcal{F}$ is identified with the opposite side modulo L.

Example 3.1.5. Let $L = \mathbb{Z}[i] = \langle i, 1 \rangle$. A fundamental domain for $\mathbb{C}/\mathbb{Z}[i]$ is given by $\mathcal{F} = \{\lambda i + \mu : 0 \leq \lambda, \ \mu < 1\}$, which is just a square (only two sides are actually included in $\mathcal{F}$). Notice that $\lambda i \equiv \lambda i + 1 \bmod L$ for all $\lambda \in \mathbb{R}$ (because $(\lambda i + 1) - \lambda i = 1 \in L$), and $\mu \equiv \mu + i \bmod L$ for all $\mu \in \mathbb{R}$ (because $i \in L$). Therefore, each side

of the square $\mathcal{F}$ is identified with the opposite side modulo the lattice L. Thus, $\mathbb{C}/\mathbb{Z}[i]$ is indeed a torus when considered as a surface. ■

Proposition 3.1.6. *Let $L = \langle w_1, w_2\rangle$ and $L' = \langle w_1', w_2'\rangle$ be lattices with oriented bases (i.e., w_1/w_2 and $w_1'/w_2' \in \mathbb{H}$).*

(1) *$L = L'$ if and only if there is a matrix $M \in \mathrm{SL}(2,\mathbb{Z})$ such that $\binom{w_1'}{w_2'} = M\binom{w_1}{w_2}$.*
(2) *There is a complex analytic (i.e., holomorphic) isomorphism of the quotients $\mathbb{C}/L$ and $\mathbb{C}/L'$ (as additive groups) if and only if $L' = \alpha L$ for some $\alpha \in \mathbb{C}$.*

In Appendix B the reader can find the definition of analytic function (Definition B.2.4). Moreover, we have also included a section that describes what it means for a map $\mathbb{C}/L \to \mathbb{C}/L'$ to be analytic (see Section B.6).

Corollary 3.1.7. *Let $L = \langle w_1, w_2\rangle$ and $L' = \langle w_1', w_2'\rangle$ be oriented bases of lattices such that there is an analytic isomorphism $\mathbb{C}/L \cong \mathbb{C}/L'$ of abelian groups. Then, there is an $\alpha \in \mathbb{C}^\times$ and $M \in \mathrm{SL}(2,\mathbb{Z})$ such that $\binom{w_1'}{w_2'} = \alpha M\binom{w_1}{w_2}$.*

The proofs of (most of) the proposition and corollary are left as exercises (Exercises 3.7.2 and 3.7.3). We will, however, need to rely on the following fact from complex analysis without giving a proof: if a map $\psi : \mathbb{C}/L \to \mathbb{C}/L'$ is an analytic isomorphism, then there is $\alpha \in \mathbb{C}^\times$ such that $L' = \alpha L$ and $\psi(z \bmod L) = \alpha z \bmod L'$. See **[Sil86]**, Ch. VI, Theorem 4.1 for more details.

Remark 3.1.8. Let $L = \langle w_1, w_2\rangle$ and $L' = \langle w_1', w_2'\rangle$ be two arbitrary lattices. Then, the map $\psi : \mathbb{C}/L \to \mathbb{C}/L'$, given by

$$\psi(\lambda w_1 + \mu w_2 \bmod L) = \lambda w_1' + \mu w_2' \bmod L'$$

for any $0 \le \lambda, \mu < 1$, is a *bijection* of sets (indeed, ψ is a bijection between the fundamental domains of $\mathbb{C}/L$ and $\mathbb{C}/L'$). In fact, ψ is also an isomorphism of abelian groups. However, in general, this map *is not analytic*.

Example 3.1.9. Let $L = \mathbb{Z}[i] = \langle i, 1\rangle$ with $w_1 = i$ and $w_2 = 1$. Let

$$M = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}).$$

Put $\binom{w_1'}{w_2'} = M\binom{w_1}{w_2} = M\binom{i}{1} = \binom{3i+5}{i+2}$. Then $\mathbb{Z}[i] = \langle i, 1\rangle = \langle 5+3i, 2+i\rangle$. Indeed, it is clear that $5+3i,\ 2+i \in \langle i, 1\rangle$. Moreover,

$$3\cdot(2+i) - (5+3i) = 1 \quad \text{and} \quad 2\cdot(5+3i) - 5\cdot(2+i) = i;$$

therefore $\langle i, 1\rangle \subseteq \langle 5+3i, 2+i\rangle$, and so they are equal lattices. Now, define $L' = \langle \frac{i}{5} + \frac{13}{5}, 1\rangle = \frac{1}{2+i}\langle 5+3i, 2+i\rangle = \frac{1}{2+i}\mathbb{Z}[i]$. By Proposition 3.1.6, there is an isomorphism $\mathbb{C}/\langle i, 1\rangle \cong \mathbb{C}/\langle \frac{i}{5} + \frac{13}{5}, 1\rangle$. ■

Suppose that $L = \langle w_1, w_2\rangle$ is an arbitrary lattice, with an oriented basis (i.e., $w_1/w_2 \in \mathbb{H}$). Then $L' = \langle \tau, 1\rangle$, with $\tau = w_1/w_2 \in \mathbb{H}$, is another lattice such that, by Prop. 3.1.6, $\mathbb{C}/L \cong \mathbb{C}/L'$. Therefore, this shows that for every lattice L, there is a lattice of the form $L' = \langle \tau, 1\rangle$ with $\tau \in \mathbb{H}$ such that $\mathbb{C}/L \cong \mathbb{C}/L'$.

When is $\mathbb{C}/\langle \tau, 1\rangle \cong \mathbb{C}/\langle \tau', 1\rangle$? If the two quotients are isomorphic, then Cor. 3.1.7 implies that there must be a matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$$

and $\alpha \in \mathbb{C}^\times$ such that

$$\begin{pmatrix} \tau' \\ 1 \end{pmatrix} = \alpha \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \tau \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha(a\tau + b) \\ \alpha(c\tau + d) \end{pmatrix}.$$

Thus, $1 = \alpha(c\tau + d)$ and so $\alpha = (c\tau + d)^{-1}$. Hence,

$$\tau' = \frac{a\tau + b}{c\tau + d} \quad \text{with} \quad ad - bc = 1.$$

If $M = (a, b; c, d)$ is a matrix in $\mathrm{SL}(2, \mathbb{Z})$ and $\tau \in \mathbb{H}$, we will write $M\tau := \frac{a\tau+b}{c\tau+d} \in \mathbb{H}$. We record our findings in the form of a proposition.

Proposition 3.1.10. *Let $L = \langle w_1, w_2\rangle$ be a lattice in $\mathbb{C}$.*

(1) *There is a $\tau \in \mathbb{H}$ such that $\mathbb{C}/L \cong \mathbb{C}/\langle \tau, 1\rangle$.*

(2) *Let $\tau, \tau' \in \mathbb{H}$. Then $\mathbb{C}/\langle \tau, 1\rangle \cong \mathbb{C}/\langle \tau', 1\rangle$ if and only if there is a matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$ such that*

$$\tau' = M\tau = \frac{a\tau + b}{c\tau + d}.$$

3.2. Functions on lattices and elliptic functions

In this section we discuss functions on $\mathbb{C}/L$. One way to construct a function $f : \mathbb{C}/L \to \mathbb{C}$ is to find a function $\widehat{f} : \mathbb{C} \to \mathbb{C}$ that is periodic with respect to the lattice L, i.e., $\widehat{f}(z+w) = \widehat{f}(z)$ for all $w \in L$. Thus, $\widehat{f}$ induces a well-defined function on $\mathbb{C}/L$ because, if $z_1 \equiv z_2 \bmod L$ (i.e., $z_2 = z_1 + w$ for some $w \in L$), then $\widehat{f}(z_1) = \widehat{f}(z_2)$. Hence, we can define $f(z \bmod L) := \widehat{f}(z)$ and this is a well-defined function on $\mathbb{C}/L$. The functions of this type are called elliptic functions.

Definition 3.2.1. An *elliptic function* (relative to a lattice $L \subset \mathbb{C}$) is a meromorphic function $f(z) : \mathbb{C} \to \mathbb{C} \cup \{\infty\}$ that satisfies $f(z+w) = f(z)$ for all $z \in \mathbb{C}$ and all $w \in L$. The set of all elliptic functions for L is denoted by $\mathbb{C}(L)$.

The most important example of an elliptic function is the Weierstrass $\wp$-function.

Definition 3.2.2. Let L be a lattice. The Weierstrass $\wp$-function relative to L is the function

$$\wp(z, L) = \frac{1}{z^2} + \sum_{0 \neq w \in L} \left(\frac{1}{(z-w)^2} - \frac{1}{w^2} \right).$$

The Weierstrass $\wp$-function is a meromorphic function on $\mathbb{C}$ and it has (double) poles at each lattice point $w \in L$. And, most importantly for us, the Weierstrass $\wp$-function is an elliptic function for the lattice L since, clearly, $\wp(z, L) = \wp(z+v, L)$ for any $v \in L$ (check this!). The Laurent series of the $\wp$-function is also very important. In order to be able to write down the Laurent series, we need to define another very important function of lattices: the Eisenstein series.

Definition 3.2.3. Let $k \geq 2$ and let L be a lattice. The Eisenstein series of L of weight $2k$ is the series

$$G_{2k}(L) = \sum_{0 \neq w \in L} \frac{1}{w^{2k}}.$$

Here, we will not worry too much about convergence, but the worried reader may be relieved to know that $G_{2k}(L)$ is absolutely convergent for $k > 1$ and $\wp(z, L)$ converges uniformly on every compact subset of $\mathbb{C} - L$ (the *worried* reader can find a proof of the

convergence in [**Sil86**], Ch. VI, Theorem 3.1). We are now ready to write down the Laurent series about $z = 0$ for the function $\wp(z, L)$.

Theorem 3.2.4. *Let L be a lattice.*

(1) *The Laurent series for $\wp(z, L)$ about $z = 0$ is given by*

$$\wp(z, L) = \frac{1}{z^2} + \sum_{k=1}^{\infty} (2k + 1) G_{2k+2}(L) z^{2k},$$

where $G_{2k+2}(L)$ is the Eisenstein series for L of weight $2k + 2$.

(2) *Let $\wp'(z, L)$ be the derivative of $\wp$ with respect to z. For all $z \in \mathbb{C}$ and $z \notin L$,*

$$\left(\frac{\wp'(z, L)}{2} \right)^2 = \wp(z, L)^3 - 15 G_4(L) \wp(z, L) - 35 G_6(L).$$

In other words, $(\wp(z, L), \frac{\wp'(z,L)}{2})$ is a point on the elliptic curve $E_L(\mathbb{C})$, where

$$E_L/\mathbb{C} : y^2 = x^3 - 15 G_4(L) x - 35 G_6(L).$$

See Exercise 3.7.4 for a proof of the first part of the theorem (part (2) is shown in [**Sil86**], Theorem VI.3.5). Theorem 3.2.4 shows that there is a map:

$$\Phi : \mathbb{C}/L \to E_L(\mathbb{C}), \quad z \bmod L \mapsto \left(\wp(z, L), \frac{\wp'(z, L)}{2} \right). \tag{3.1}$$

It turns out that the map Φ has all the "nice" properties that one would hope for: it is a complex analytic isomorphism of abelian groups. Moreover, if $E/\mathbb{C} : y^2 = x^3 + Ax + B$ is an elliptic curve, then there is a lattice $L \subset \mathbb{C}$ such that $\Phi : \mathbb{C}/L \cong E(\mathbb{C})$. This result is usually called the *uniformization theorem*:

Theorem 3.2.5. (Uniformization theorem)

(1) *Let L be a lattice. Then the equation $y^2 = x^3 - 15 G_4(L) x - 35 G_6(L)$ is non-singular (i.e., its discriminant is $\neq 0$) and defines an elliptic curve $E_L/\mathbb{C}$. Moreover, the map $\Phi : \mathbb{C}/L \to E_L(\mathbb{C})$ defined in* Eq. (3.1) *is a complex analytic isomorphism of abelian groups.*

(2) *Let $E/\mathbb{C} : y^2 = x^3 + Ax + B$ be an elliptic curve. Then there exists a lattice $L \subset \mathbb{C}$ such that $A = -15G_4(L)$, $B = -35G_6(L)$ and $\mathbb{C}/L$ is isomorphic to $E(\mathbb{C})$ via Φ.*

For a proof of the uniformization theorem, see **[DS05]**, §1.4.

Example 3.2.6. Let $E/\mathbb{Q} : y^2 = x^3 - x$. The lattice that corresponds to this elliptic curve is

$$L = \langle\ (13.5823633497\ldots)i,\ 13.5823633497\ldots\ \rangle$$

because $-15G_4(L) = -1$ and $-35G_6(L) = 0$. Let us define a quantity $\Omega_E = 13.5823633497\ldots$. Then, $L = \Omega_E \cdot \langle i, 1\rangle$ and, therefore (by Prop. 3.1.6), $E(\mathbb{C}) \cong \mathbb{C}/\langle i, 1\rangle$. ■

3.3. Elliptic curves and the upper half-plane

The uniformization theorem tells us that every lattice L determines an elliptic curve $E_L/\mathbb{C}$ and, conversely, for every elliptic curve $E/\mathbb{C}$ there is a lattice L that produces E, i.e., $E(\mathbb{C}) \cong \mathbb{C}/L$. Proposition 3.1.10 tells us that we can find a lattice of the form $\langle\tau, 1\rangle$, with $\tau \in \mathbb{H}$, such that $E(\mathbb{C}) \cong \mathbb{C}/\langle\tau, 1\rangle$. Thus, every τ in the complex upper half-plane determines a lattice $L_\tau = \langle\tau, 1\rangle$, which in turn determines a $\mathbb{C}$-isomorphism class of an elliptic curve $E(\mathbb{C}) \cong \mathbb{C}/\langle\tau, 1\rangle$. The choice of τ, however, is not unique. Remember that, also by Prop. 3.1.10, if τ' is another element of the complex upper half-plane and there exists a matrix $M \in \mathrm{SL}(2, \mathbb{Z})$ such that $\tau' = M\tau$ $(= \frac{a\tau+b}{c\tau+d})$, then $E(\mathbb{C}) \cong \mathbb{C}/\langle\tau', 1\rangle$.

The discussion in the preceding paragraph motivates the definition of an equivalence relation between points in $\mathbb{H}$ modulo $\mathrm{SL}(2, \mathbb{Z})$. For the sake of brevity, we will write $\Gamma(1)$ for $\mathrm{SL}(2, \mathbb{Z})$, and we will call it *the modular group*. Later on, we shall describe other subgroups $\Gamma(N)$ that will justify this notation (see Definition 3.5.1).

Definition 3.3.1. We say that two points $\tau, \tau' \in \mathbb{H}$ are *equivalent relative to the modular group* $\Gamma(1)$ if there is a matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$$

such that $\tau' = M\tau$. This defines an equivalence relation (see Exercise 3.7.6), and the set of all equivalence classes is denoted by $Y(1) = \mathbb{H}/\Gamma(1)$:

$$\begin{aligned} Y(1) &= \mathbb{H}/\Gamma(1) \\ &= \frac{\{z = a + bi \in \mathbb{C} : b > 0\}}{\{z \sim z' \text{ if and only if } z' = Mz \text{ for some } M \in \mathrm{SL}(2,\mathbb{Z})\}}. \end{aligned}$$

Remark 3.3.2. The reader should also notice that, for any $\tau \in \mathbb{H}$, the matrices $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $-M = \begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$ afford the same action on τ. Indeed,

$$(-M)\tau = \frac{-a\tau - b}{-c\tau - d} = \frac{a\tau + b}{c\tau + d} = M\tau.$$

Thus, sometimes the equivalence relation is defined with respect to the quotient $\mathrm{SL}(2,\mathbb{Z})/\{\pm \mathrm{Id}\}$.

Example 3.3.3. For instance, $z = 1 + i$ and $z' = 7 + i$ are representatives of the same equivalence class in $Y(1) = \mathbb{H}/\Gamma(1)$ because

$$\begin{aligned} M' &= \begin{pmatrix} 1 & 6 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}), \text{ and} \\ Mz &= M \cdot (1+i) = \frac{1 \cdot (1+i) + 6}{0 \cdot (1+i) + 1} = 7 + i = z'. \end{aligned}$$

Similarly, $z = 1+i$ and $z'' = \frac{i+27}{10}$ are representatives of the same class in $Y(1)$. In this case, the transitional matrix is $M'' = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$ and, indeed, $z'' = Mz$ (check this). Proposition 3.1.10 implies that the elliptic curves that correspond to the quotients $\mathbb{C}/\langle z, 1\rangle$, $\mathbb{C}/\langle z', 1\rangle$ and $\mathbb{C}/\langle z'', 1\rangle$ are all isomorphic (over $\mathbb{C}$). ■

As in the case of the quotient $\mathbb{C}/L$ by a lattice L (see Definition 3.1.4), we would like to find a fundamental domain for the quotient $\mathbb{H}/\Gamma(1)$.

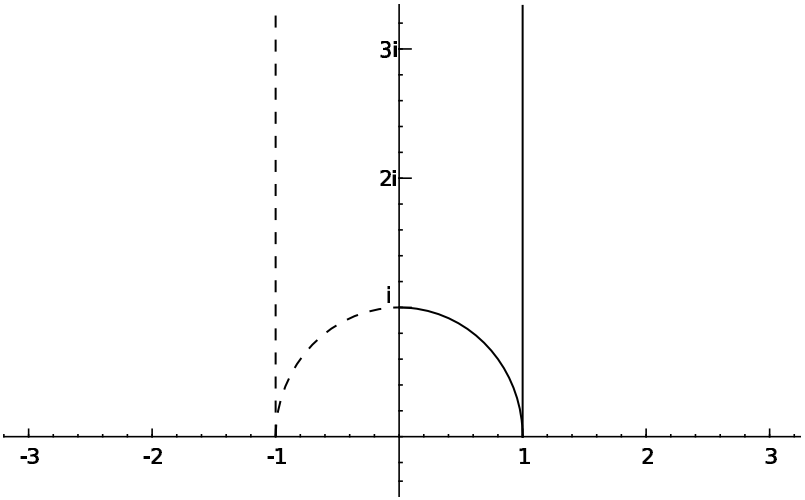


Figure 3. The fundamental domain $\mathcal{F}(1)$ for the quotient $\mathbb{H}/\Gamma(1)$.

Proposition 3.3.4. *Let $\mathcal{F}(1) \subset \mathbb{C}$ be the following set of complex numbers:*

$$\begin{aligned}\mathcal{F}(1) &= \left\{ z = a + bi \in \mathbb{C} : |z| > 1 \text{ and } -\frac{1}{2} < a = \Re(z) \leq \frac{1}{2} \right\} \\ &\cup \left\{ z = a + bi \in \mathbb{C} : |z| = 1 \text{ and } 0 \leq a = \Re(z) \leq \frac{1}{2} \right\}.\end{aligned}$$

Then $\mathcal{F}(1)$ is a fundamental domain for $\mathbb{H}/\Gamma(1)$, i.e.,

(1) *If $w \in \mathbb{H}$, then there is $z \in \mathcal{F}(1)$ such that $w \sim z$ in $\mathbb{H}/\Gamma(1)$; i.e., there is $M \in \mathrm{SL}(2,\mathbb{Z})$ such that $w = Mz$, and*

(2) *If z, z' are two distinct elements of $\mathcal{F}(1)$, then $z \not\sim z'$ in $\mathbb{H}/\Gamma(1)$; that is, the equivalence classes of z and z' are distinct.*

The proof of part (1) of Proposition 3.3.4 is left as an exercise (Exercises 3.7.1 and 3.7.7). The proof of (2) can be found, for example, in [**Ser77**], Ch. VII. In the proof of (1), in Exercise 3.7.7, we

found two distinguished matrices of $\mathrm{SL}(2,\mathbb{Z})$:

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

The action of the matrices S and T on $\tau \in \mathbb{H}$ is simply $T\tau = \tau + 1$ and $S\tau = -\frac{1}{\tau}$. Notice also that $S^2 = \mathrm{Id}$ and $(ST)^3 = -\mathrm{Id}$. As a corollary of Prop. 3.3.4, we show that the subgroup generated by S and T (i.e., the group G of Exercise 3.7.7) is all of $\mathrm{SL}(2,\mathbb{Z})$.

Corollary 3.3.5. *The modular group $\Gamma(1) = \mathrm{SL}(2,\mathbb{Z})$ is generated by the matrices S and T.*

Proof. Let M be a matrix in $\mathrm{SL}(2,\mathbb{Z})$ and let $\tau \in \mathbb{H}$ be a fixed complex number in the interior of $\mathcal{F}(1)$ (e.g., $\tau = 2i$ works). Let $\tau' = M\tau$ and write G for the subgroup of $\mathrm{SL}(2,\mathbb{Z})$ generated by S and T. Notice that $-\mathrm{Id} = (ST)^3 \in G$. We want to show that $G = \Gamma(1)$.

Exercise 3.7.7 says that there exists a matrix $M' \in G$ such that $\tau'' = M'\tau'$ lies in $\mathcal{F}(1)$. Thus, $\tau'' = M'\tau' = (M'M)\tau$. Since M and M' belong to $\mathrm{SL}(2,\mathbb{Z})$, their product $M'M \in \mathrm{SL}(2,\mathbb{Z})$ and therefore $\tau'' \sim \tau$ in $\mathbb{H}/\Gamma(1)$ by definition. Moreover, both τ and τ'' are in $\mathcal{F}(1)$. But Prop. 3.3.4, part (2), says that this is impossible unless $\tau = \tau''$. Hence, $\tau = M'M\tau$ and this implies that $M'M = \pm\mathrm{Id}$ (here we are using the fact that τ is in the interior of $\mathcal{F}(1)$). Thus, $M = \pm(M')^{-1} \in G$. Since $M \in \mathrm{SL}(2,\mathbb{Z})$ was arbitrary, we conclude that $\Gamma(1) \leq G$. The reverse inclusion is obvious. Thus, $G = \Gamma(1)$, as claimed. ∎

3.4. The modular curve $X(1)$

Proposition 3.3.4 shows that each point on the fundamental domain $\mathcal{F}(1)$ represents a unique class in the quotient $Y(1) = \mathbb{H}/\Gamma(1)$, and every class of $Y(1)$ has a representative in $\mathcal{F}(1)$. If we considered $\mathbb{H}/\Gamma(1)$ as a surface, then it would be homeomorphic to a sphere with one point missing (to see this, identify the sides in the boundary of $\mathcal{F}(1)$). In order to *compactify* $Y(1)$, we add one point, a point at infinity, that makes $Y(1) \cup \{\infty\}$ into a compact surface (a sphere), denoted $X(1)$. In the interest of space, time and ink, we will not discuss the

topology of $Y(1)$ and $X(1)$ (for instance, see **[Sil94]**, Ch. 1, §2 or see **[DS05]**, Sections 2.1 and 2.2 for a more generalized approach). The formal construction of this point at infinity is the following.

We define an extended upper half plane by $\mathbb{H}^* := \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$, i.e., the union of $\mathbb{H}$ and a copy of a projective line over $\mathbb{Q}$. The points in the projective line of the form $\{[s,1] : s \in \mathbb{Q}\}$ are simply the rational points of the real axis in the complex plane (so $[s,1]$ stands for $s + 0 \cdot i \in \mathbb{Q} \subset \mathbb{R}$). The remaining point of $\mathbb{P}^1(\mathbb{Q})$ is the point at infinity $\infty := [1,0]$. If the reader needs to review the basics about the projective line, see Appendix C.1.

We also extend the action of $\Gamma(1)$ to all of $\mathbb{H}^*$ as follows. Let $M = (a,b;c,d) \in \mathrm{SL}(2,\mathbb{Z})$ and let $[s,t] \in \mathbb{P}^1(\mathbb{Q})$. Then we define

$$M[s,t] = (-M)[s,t] := [as + bt, cs + dt] \in \mathbb{P}^1(\mathbb{Q}).$$

Notice that $[s,1]$, which we may identify with $s \in \mathbb{Q}$, is sent to $M[s,1] = [as + b, cs + d]$, and as long as $cs + d \neq 0$, then $M[s,1] = [\frac{as+b}{cs+d}, 1]$, so that $s \in \mathbb{Q}$ is sent to $Ms = \frac{as+b}{cs+d} \in \mathbb{Q}$. Thus, this action is clearly consistent with the previous definition of the action of $\Gamma(1)$ on $\mathbb{H}$. The point at infinity can also be treated similarly.

Example 3.4.1. Let $M = \begin{pmatrix} 5 & 6 \\ 4 & 5 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$. Let $s = [3,1] \in \mathbb{Q}$. Then

$$Ms = M[3,1] = [5 \cdot 3 + 6, 4 \cdot 3 + 5] = [21, 17],$$

or, equivalently, $Ms = M \cdot 3 = \frac{5 \cdot 3 + 6}{4 \cdot 3 + 5} = \frac{21}{17}$. One needs to be careful with zeros in the denominators! For instance, let $s' = -\frac{5}{4}$. Then:

$$Ms' = M \cdot \left(-\frac{5}{4}\right) = \frac{5 \cdot \left(-\frac{5}{4}\right) + 6}{4 \cdot \left(-\frac{5}{4}\right) + 5} = \frac{-\frac{1}{4}}{0} \text{ “=” } \infty.$$

The previous equation can be formalized using projective coordinates:

$$M[s',1] = M\left[-\frac{5}{4}, 1\right] = \left[-\frac{1}{4}, 0\right] = [1,0].$$

We can also calculate the action of M on ∞ using the usual laws of limits:

$$M \cdot \infty \text{ “=” } \frac{5 \cdot \infty + 6}{4 \cdot \infty + 5} \text{ “=” } \frac{5}{4}.$$

Once again, this calculation can be formalized using projective coordinates:

$$M[1,0] = [5 \cdot 1 + 6 \cdot 0, 4 \cdot 1 + 5 \cdot 0] = [5,4] = \left[\frac{5}{4}, 1\right].$$

■

We are ready to define $X(1)$. The definition now is identical to the definition of $Y(1)$ in Definition 3.3.1:

Definition 3.4.2. We say that two points $\tau, \tau' \in \mathbb{H}^* = \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ are equivalent relative to the modular group $\Gamma(1)$ if there is a matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$$

such that $\tau' = M\tau$. This defines an equivalence relation, and the set of all equivalence classes is denoted by $X(1) = \mathbb{H}^*/\Gamma(1)$:

$$\begin{aligned} X(1) &= \mathbb{H}^*/\Gamma(1) \\ &= \frac{\{z = a + bi \in \mathbb{C} : b > 0\} \cup \{s \in \mathbb{Q}\} \cup \{\infty\}}{\{z \sim z' \text{ if and only if } z' = Mz \text{ for some } M \in \mathrm{SL}(2,\mathbb{Z})\}}. \end{aligned}$$

At the beginning of this section we claimed that we were going to adjoin *only one point* to $Y(1)$, to compactify the surface, but it would seem we added infinitely many points (i.e., all of $\mathbb{P}^1(\mathbb{Q})$). However, all the points in $\mathbb{P}^1(\mathbb{Q})$ represent the same class in $\mathbb{H}^*/\Gamma(1)$, and therefore $X(1) = \mathbb{H}^*/\Gamma(1)$ only contains one extra point more than $Y(1)$.

Proposition 3.4.3. *Let s and s' be two elements of $\mathbb{P}^1(\mathbb{Q})$. Then there exists a matrix M in $\mathrm{SL}(2,\mathbb{Z})$ such that $s' = Ms$.*

The proof is left as an exercise (Exercise 3.7.8). Proposition 3.4.3 implies that every point in $\mathbb{P}^1(\mathbb{Q})$ is equivalent to $\infty = [1,0]$ in $\mathbb{H}^*/\Gamma(1)$. Thus, $X(1) = Y(1) \cup \{\infty\}$ as we wanted. The point ∞ is called a *cusp*.

In the next couple of sections we are going to generalize the construction of $X(1)$ and define other types of modular curves that will show up later on. First of all, we need to talk about the subgroups of $\mathrm{SL}(2,\mathbb{Z})$ that we are most interested in.

3.5. Congruence subgroups

In this section we define several types of subgroups Γ of $\mathrm{SL}(2,\mathbb{Z})$ that come up often in the theory of modular forms. Later on, we will define other modular curves as quotients $\mathbb{H}^*/\Gamma$ in the same way that we have defined $X(1)$ above.

Definition 3.5.1. Let $N \geq 1$ be an integer. We define subgroups of $\mathrm{SL}(2,\mathbb{Z})$ by

$$\begin{aligned}
\Gamma_0(N) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) : c \equiv 0 \bmod N \right\}, \\
\Gamma_1(N) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) : c \equiv 0,\ a \equiv d \equiv 1 \bmod N \right\}, \\
\Gamma(N) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) : b \equiv c \equiv 0,\ a \equiv d \equiv 1 \bmod N \right\}.
\end{aligned}$$

We say that a subgroup G of $\mathrm{SL}(2,\mathbb{Z})$ is a *congruence subgroup* if G contains $\Gamma(N)$ for some integer $N \geq 1$.

The reader should check that, indeed, $\Gamma_0(N)$, $\Gamma_1(N)$ and $\Gamma(N)$ are subgroups of $\mathrm{SL}(2,\mathbb{Z})$ for any $N \geq 1$. Also notice that, for a fixed $N \geq 1$, we have inclusions $\Gamma(N) \leq \Gamma_1(N) \leq \Gamma_0(N)$. The equality $\Gamma(1) = \mathrm{SL}(2,\mathbb{Z})$ follows from Definition 3.5.1, which explains our previous notation for the modular group.

Example 3.5.2. Let $N = 5$. The following matrices belong to $\Gamma_0(5)$:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 5 & 1 \end{pmatrix},$$
$$\begin{pmatrix} -2 & -1 \\ 5 & 2 \end{pmatrix}, \begin{pmatrix} -3 & -1 \\ 10 & 3 \end{pmatrix},$$

and, in fact, one can show that these matrices (and their inverses) generate all of $\Gamma_0(5)$. The matrices

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -49 & 23 \\ 115 & -54 \end{pmatrix}, \begin{pmatrix} 11 & 4 \\ -25 & -9 \end{pmatrix}, \begin{pmatrix} 66 & 23 \\ 175 & 61 \end{pmatrix}$$

are some examples of elements of $\Gamma_1(5)$, but they are not a complete generating set for $\Gamma_1(5)$. A complete list of generators can be found, for example, using Sage with the command `Gamma1(5).gens()`. ■

3.6. Modular curves

In this section we generalize the definition of $X(1)$, as in Definition 3.4.2, in order to define more general modular curves. To do so, we simply replace $\Gamma(1)$ by any congruence subgroup Γ defined in Section 3.5.

Let Γ be a fixed congruence subgroup. We say that two points $\tau, \tau' \in \mathbb{H}^* = \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ are equivalent relative to Γ if there is a matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$$

such that $\tau' = M\tau$. This defines an equivalence relation, and the set of all equivalence classes is denoted by $X = \mathbb{H}^*/\Gamma$:

$$X = \mathbb{H}^*/\Gamma = \frac{\{z = a + bi \in \mathbb{C} : b > 0\} \cup \{s \in \mathbb{Q}\} \cup \{\infty\}}{\{z \sim z' \text{ if and only if } z' = Mz \text{ for some } M \in \Gamma\}}.$$

The space X is called a *modular curve* (indeed, X may be viewed as a curve over $\mathbb{C}$ or as a real surface). The *cusps* of $\mathbb{H}^*/\Gamma$ are those elements in the quotient that have a representative in $\mathbb{P}^1(\mathbb{Q})$. Recall that $X(1)$ had only 1 cusp. However, other modular curves have multiple distinct cusps.

Let $N \geq 1$. The modular curves that correspond to the congruence subgroups $\Gamma_0(N)$, $\Gamma_1(N)$ and $\Gamma(N)$ are usually denoted, respectively, by $X_0(N)$, $X_1(N)$ and $X(N)$.

Example 3.6.1. Let $p \geq 2$ and let $X_0(p) = \mathbb{H}^*/\Gamma_0(p)$. Then $X_0(p)$ has exactly two cusps. The points $0 = [0, 1]$ and $\infty = [1, 0]$ are inequivalent in $X_0(p)$ and are representatives of the two non-trivial cusps. See Exercise 3.7.10. ■

Remark 3.6.2. In Proposition 3.3.4 we found $\mathcal{F}(1)$, a fundamental domain for the action of $\mathrm{SL}(2, \mathbb{Z})$ on $\mathbb{H}$. Similarly, if Γ is a congruence subgroup, one can find a fundamental domain $\mathcal{F}(\Gamma)$ for the action of Γ on $\mathbb{H}$. We write $\mathcal{F}(N)$, $\mathcal{F}_1(N)$ and $\mathcal{F}_0(N)$, respectively, for the fundamental domains for the action of $\Gamma(N)$, $\Gamma_1(N)$ and $\Gamma_0(N)$ on $\mathbb{H}$. Helena Verrill **[Ver05]** has developed a great applet to draw fundamental domains for modular curves. See Figure 4 for an example of a fundamental domain for $\Gamma_0(11)$.

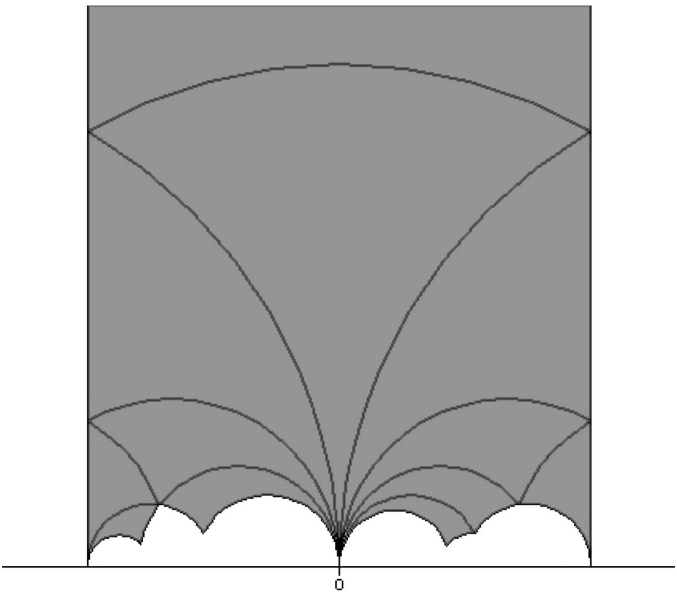


Figure 4. A fundamental domain (whole shaded region) for the action of $\Gamma_0(11)$, obtained with Verrill's applet. The domain is in fact infinitely long in the positive imaginary direction (upwards), but we had to cut the domain to be able to fit it on the page. Each hyperbolic triangle inside the shaded region is a fundamental domain for $\mathrm{SL}(2,\mathbb{Z})$.

Remark 3.6.3. As a consequence of the uniformization theorem (Thm. 3.2.5) and Prop. 3.1.10, every class $[\tau] \in X(1)$ such that τ is not a cusp (sometimes we say non-cuspidal τ) corresponds to an elliptic curve $E/\mathbb{C} \cong \mathbb{C}/\langle \tau, 1\rangle$ and, conversely, if $E/\mathbb{C}$ is an elliptic curve, there is a unique class $[\tau] \in X(1)$ such that $E/\mathbb{C} \cong \mathbb{C}/\langle \tau, 1\rangle$. Thus, the non-cuspidal points on $X(1)$ classify elliptic curves up to isomorphism over $\mathbb{C}$.

Similarly, one can show that the modular curves $X_0(N)$, $X_1(N)$ and $X(N)$ have interpretations in terms of elliptic curves together with some extra data. For instance, $X_0(N)$ classifies pairs (E, C) of elliptic curves E with a fixed subgroup $C \subseteq E(\mathbb{C})$ of order N up to isomorphism over $\mathbb{C}$. The curve $X_1(N)$ classifies pairs (E, P) of

elliptic curves E with a fixed point $P \in E(\mathbb{C})$ of exact order N up to isomorphism over $\mathbb{C}$.

Remark 3.6.4. One aspect of modular curves that is not at all obvious is the fact that modular curves have algebraic models; i.e., if Γ is a congruence subgroup, then $\mathbb{H}^*/\Gamma$ is a compact Riemann surface and it has a model as a projective algebraic curve over $\mathbb{C}$, given by polynomial equations. The modular curves for $\Gamma_0(N)$ have the surprising property that they have a canonical model defined over $\mathbb{Q}$. The reason is that the modular j-invariant function $j(z)$ (see Example 4.1.10) and the function $j(Nz)$ satisfy an algebraic equation $F_N(j(z), j(Nz)) = 0$, with $F_N(x, y) \in \mathbb{Q}[x, y]$, which gives an algebraic model for $X_0(N)$. However, this is typically a *highly singular* model, which can be transformed into a non-singular model for the modular curve. For instance,

(1) The curve $X_0(11) = \mathbb{H}^*/\Gamma_0(11)$ has a model $y^2 + y = x^3 - x^2 - 10x - 20$ (notice that it is an elliptic curve!).

(2) The curve $X_1(11) = \mathbb{H}^*/\Gamma_1(11)$ has a model $y^2+y = x^3-x^2$.

(3) The curve $X_0(14)$ has a model $y^2 + xy + y = x^3 + 4x - 6$.

(4) The curve $X_1(14)$ has a model $y^2 + xy + y = x^3 - x$.

(5) The curve $X_1(13)$ has a model $y^2 + (x^3 - x^2 - 1)y = x^2 - x$. This *is not* an elliptic curve (it has genus 2, not 1). The examples above (1)-(4) are nice but the model of a modular curve will be often much more complicated than a cubic.

We conclude this chapter with some genus formulas for $X_0(N)$, due to Ogg, Shimura, and others. The genus can be calculated using the Hurwitz genus formula and the ramification points of the quotient map $X_0(N) \to X(1)$.

Theorem 3.6.5. *Let $N \geq 1$ be an integer and let $X_0(N)$ be the modular curve $\mathbb{H}^*/\Gamma_0(N)$. Let g be the genus of the curve $X_0(N)$. Then:*

$$g = 0 \quad \text{if} \quad N = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 16, 18, 25;$$

$$g = 1 \quad \text{if} \quad N = 11, 14, 15, 17, 19, 20, 21, 24, 27, 32, 36, 49;$$

$$g = 2 \quad \text{if} \quad N = 22, 23, 26, 28, 29, 31, 37, 50.$$

Moreover, if $p > 3$ is prime, then:

$$\begin{aligned}
\text{genus}(X_0(p)) &= \begin{cases} \left[\frac{p+1}{12}\right] - 1 & \text{if } p \equiv 1 \bmod 12; \\ \left[\frac{p+1}{12}\right] & \text{otherwise,} \end{cases} \\
\text{genus}(X_1(p)) &= 1 + \frac{(p-1)(p-11)}{24}, \quad \text{and} \\
\text{genus}(X(p)) &= 1 + \frac{(p^2-1)(p-6)}{24},
\end{aligned}$$

where $[x]$ is the greatest integer $\leq x$.

The genus formulas for $X(p)$, $X_0(p)$ and $X_1(p)$ are consequences of the Hurwitz and Riemann-Hurwitz genus formulas (see Exercises 3.1.4, 3.1.5 and 3.1.6 of **[DS05]** or see Chapter 1 of **[Shi73]** for proofs). The list of all modular curves $X_0(N)$ with genus 0, 1 or 2 can be found in **[Maz72]**. For more general genus formulas see **[DS05]**, Section 3.1.

3.7. Exercises

Exercise 3.7.1. Let $a, b, c, d \in \mathbb{R}$, $\tau \in \mathbb{C}$ and $\tau \notin \mathbb{R}$. Show that:

(1) The imaginary part of $\tau' = \frac{a\tau+b}{c\tau+d}$ is $\text{Im}(\tau') = \frac{(ad-bc)\,\text{Im}(\tau)}{|c\tau+d|^2}$.

(2) If $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}(2,\mathbb{Z})$ and $\tau \in \mathbb{H}$, then $M\tau \in \mathbb{H}$.

Exercise 3.7.2. In this exercise we study the relationships between different bases of a lattice.

(1) Let $L = \langle i, 1\rangle$ be the lattice of Gaussian integers $\mathbb{Z}[i]$. Let a, b, c, d be integers such that $ad - bc = 1$. Show that the lattice L' generated by $w_1 = ai + b$ and $w_2 = ci + d$ is also $\mathbb{Z}[i]$.

(2) More generally, let L be a lattice generated by w_1 and $w_2 \in \mathbb{C}$ with $w_1/w_2 \in \mathbb{H}$. Let

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be a matrix in $\text{SL}(2,\mathbb{Z})$, i.e., $a, b, c, d \in \mathbb{Z}$ and $ad - bc = 1$. Let $\binom{w_1'}{w_2'} = M\binom{w_1}{w_2}$, where the operation here is the usual matrix multiplication of vectors, i.e., $w_1' = aw_1 + bw_2$ and

$w_2' = cw_1 + dw_2$. Show that $w_1'/w_2' \in \mathbb{H}$ and the lattice generated by w_1' and w_2' is also L. (Hint: do Exercise 3.7.1. Also, notice that M is an invertible matrix.)

(3) Conversely, suppose that $L = \langle w_1, w_2\rangle = \langle w_1', w_2'\rangle$, for some $w_i, w_i' \in \mathbb{C}$, such that w_1/w_2, $w_1'/w_2' \in \mathbb{H}$. Show that there is a matrix $M \in \mathrm{SL}(2,\mathbb{Z})$ such that $\binom{w_1'}{w_2'} = M\binom{w_1}{w_2}$.

Exercise 3.7.3. Let L and L' be lattices in $\mathbb{C}$. Let $\alpha \in \mathbb{C}^\times$ and suppose that $L = \alpha L'$. Show that the map $\psi : \mathbb{C}/L \to \mathbb{C}/L'$ defined by $\psi(z \bmod L) = \alpha z \bmod L'$ is an analytic map and it is also an isomorphism of abelian groups.

Exercise 3.7.4. This exercise shows part (a) of Theorem 3.2.4.

(a) Find the Taylor series of $f(x) = \frac{1}{(1-x)^2}$ centered at $x = 0$.

(b) Use (a) to find the Laurent series of $\wp(z, L)$ centered around $z = 0$. Hint:

$$\frac{1}{(z-w)^2} - \frac{1}{w^2} = \frac{1}{w^2}\left(\frac{1}{(1-\frac{z}{w})^2} - 1\right).$$

Exercise 3.7.5. Let $E/\mathbb{C}$ be an elliptic curve. Show that $E[m] \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$. (Hint: use the uniformization theorem, Thm. 3.2.5. What is the m-torsion of $\mathbb{C}/L$?)

Exercise 3.7.6. The goal of this problem is to show that the relation that appears in Definition 3.3.1 is indeed an equivalence relation. Let $M = (a, b; c, d) \in \mathrm{SL}(2,\mathbb{Z})$, $\tau, \tau' \in \mathbb{H}$ and define $M\tau = \frac{a\tau+b}{c\tau+d}$. We say that $\tau \sim \tau'$ if there is a matrix $M \in \mathrm{SL}(2,\mathbb{Z})$ such that $\tau' = M\tau$. Show that:

(1) (Reflexive) $\tau \sim \tau$ for all $\tau \in \mathbb{H}$;

(2) (Symmetric) if $\tau \sim \tau'$, then $\tau' \sim \tau$ for all $\tau, \tau' \in \mathbb{H}$;

(3) (Transitive) if $\tau \sim \tau'$ and $\tau' \sim \tau''$, then $\tau \sim \tau''$ for all $\tau, \tau, \tau'' \in \mathbb{H}$.

Exercise 3.7.7. Let G be the subgroup of $\mathrm{SL}(2,\mathbb{Z})$ generated by the matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

In other words, G is the group of all matrices that can be obtained as "words" in the letters S, T, and T^{-1} (e.g. $M = S \cdot T \cdot S \cdot T^3 \cdot S \in G$). The goal of this exercise is to show that for all $\tau \in \mathbb{H}$ there is $M \in G$ such that $M\tau$ is in the fundamental domain $\mathcal{F}(1)$ defined in Prop. 3.3.4.

(1) Let $\tau \in \mathbb{H}$ be fixed. The set

$$U = \{\mathrm{Im}(M\tau) : M \in G\} \subset \mathbb{R}^{>0}$$

has a maximum element, i.e., there is $M_0 \in G$ such that $\mathrm{Im}(M_0\tau)$ is the maximum element of U. (Hint: show that $|c\tau + d| \to \infty$ as $|c| + |d| \to \infty$. Then use Prob. 3.7.1.)

(2) Let τ and M_0 be as in (1). Show that there is $n \in \mathbb{Z}$ such that

$$|\Re(T^n M_0\tau)| \leq \frac{1}{2}.$$

(3) Let τ, M_0 and n be as above. Show that if $|T^n M_0\tau| < 1$, then

$$\mathrm{Im}(ST^n M_0\tau) > \mathrm{M_0}\tau$$

contradicting the definition of M_0. Hence $|T^n M_0\tau| \geq 1$.

(4) If $\tau \in \mathcal{F}'$ with

$$\begin{aligned}\mathcal{F}' &= \left\{z = a + bi \in \mathbb{C} : |z| > 1 \text{ and } a = \Re(z) = -\frac{1}{2}\right\} \\ &\cup \left\{z = a + bi \in \mathbb{C} : |z| = 1 \text{ and } -\frac{1}{2} \leq a = \Re(z) < 0\right\},\end{aligned}$$

then there is $M \in G$ such that $M\tau \in \mathcal{F}(1)$.

(5) Conclude that for every $\tau \in \mathbb{H}$ there is $M \in G$ such that $M\tau \in \mathcal{F}(1)$.

Exercise 3.7.8. Prove Proposition 3.4.3. In particular, show that ∞ is equivalent to all rational points $[s, 1] \in \mathbb{P}^1(\mathbb{Q})$, with $s \in \mathbb{Q}$, in $X(1)$.

Exercise 3.7.9. Let $N > 1$ be fixed. Prove that $\Gamma(N)$, $\Gamma_1(N)$ and $\Gamma_0(N)$ are subgroups of $\mathrm{SL}(2, \mathbb{Z})$, and $\Gamma(N) \leq \Gamma_1(N) \leq \Gamma_0(N)$.

Exercise 3.7.10. Let p be a prime and let $X_0(p) = \mathbb{H}^*/\Gamma_0(p)$. Show that $X_0(p)$ has exactly two cusps. In particular, show that if $[s, t] \in$

$\mathbb{P}^1(\mathbb{Q})$, then either there is a matrix $M \in \Gamma_0(p)$ such that $M[s,t] = [0,1]$ or there is a matrix $M' \in \Gamma_0(p)$ such that $M'[s,t] = [1,0]$, but both cannot occur simultaneously (i.e., $0 \not\sim \infty$ in $X_0(p)$).

Chapter 4

Modular forms

In this chapter we define modular forms as functions on modular curves (see Definition 3.4.2 and Section 3.6). For further introductory reading on modular forms (and modular curves), we refer the reader to [**Ser77**], Ch. VII, and the first two chapters of [**DS05**]. See Appendix A.2 for a brief introduction to computing spaces of modular forms using Sage, and see [**Ste07**] for a thorough treatment.

4.1. Modular forms for the modular group

First, we define meromorphic functions on a modular curve $\mathbb{H}^*/\Gamma$. For the definition of congruence subgroup, see Section 3.5.

Definition 4.1.1. Let Γ be a congruence subgroup of $\mathrm{SL}(2,\mathbb{Z})$ and let $X(\Gamma)$ be the modular curve $\mathbb{H}^*/\Gamma$. A meromorphic function $f : \mathbb{H}^*/\Gamma \to \mathbb{C} \cup \{\infty\}$ is called *a modular function* for Γ. In other words, a modular function is a function $f : \mathbb{H}^* \to \mathbb{C} \cup \{\infty\}$ such that:

(1) f is meromorphic on $\mathbb{H}^*$, so it is meromorphic at all points on $\mathbb{H}$ and all points on $\mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}$; and

(2) $f(z) = f(Mz)$ for any $M \in \Gamma$; i.e., $f(z) = f(\frac{az+b}{cz+d})$, where $M = (a, b; c, d)$ is a matrix in Γ.

Remark 4.1.2. Let $\Gamma = \mathrm{SL}(2,\mathbb{Z})$ and let $X(1) = \mathbb{H}^*/\mathrm{SL}(2,\mathbb{Z})$. Then a modular function for $\mathrm{SL}(2,\mathbb{Z})$ is a function $f : \mathbb{H}^* \to \mathbb{C} \cup \{\infty\}$ that

is meromorphic on all of $\mathbb{H}^*$ and such that $f(z) = f(Mz)$ for all $M \in \mathrm{SL}(2,\mathbb{Z})$. In particular, $f(z) = f(Sz) = f(-1/z)$ and $f(z) = f(Tz) = f(z+1)$. In fact, if $f(z) = f(Mz)$ for $M = S$ and T, then it also holds for all $M \in \mathrm{SL}(2,\mathbb{Z})$, because S and T generate all of $\mathrm{SL}(2,\mathbb{Z})$ by Corollary 3.3.5.

It turns out that the conditions on the definition of modular function are quite restrictive and, as a result, there are very few interesting examples. For instance, the modular functions for $\mathrm{SL}(2,\mathbb{Z})$ are just $\mathbb{C}(j)$, where $j(z)$ is the modular j-invariant (see Example 4.1.10). We extend the definition a bit. We begin with the definition of modular forms for $\mathrm{SL}(2,\mathbb{Z})$.

Definition 4.1.3. A function $f : \mathbb{H} \to \mathbb{C} \cup \{\infty\}$ is *weakly modular of weight* k for $\mathrm{SL}(2,\mathbb{Z})$ if:

(1) f is meromorphic on $\mathbb{H}$; and

(2) $f(Mz) = (cz+d)^k f(z)$ for any $M \in \mathrm{SL}(2,\mathbb{Z})$; i.e., $f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k f(z)$, where M is a matrix $(a,b;c,d) \in \mathrm{SL}(2,\mathbb{Z})$.

Since the matrix T is an element of $\mathrm{SL}(2,\mathbb{Z})$, if f is weakly modular, then

$$f(z) = f(Tz) = f(z+1) \quad \text{for all} \quad z \in \mathbb{H}.$$

This periodicity of f means that, if we set $q = e^{2\pi i z}$, then we can express f as a Laurent series:

$$f(z) = \sum_{n=-\infty}^{\infty} a_n q^n = \cdots + \frac{a_{-2}}{q^2} + \frac{a_{-1}}{q} + a_0 + a_1 q + a_2 q^2 + \cdots,$$

where a_n are called the Fourier coefficients of f.

(a) We say that f is a *modular function of weight* k for $\mathrm{SL}(2,\mathbb{Z})$ if f satisfies (1) and (2) above and f is *meromorphic at the cusp* ∞ of the modular curve $X(1) = \mathbb{H}^*/\mathrm{SL}(2,\mathbb{Z})$. This means that the Laurent expansion of $f(z)$ must be of the form

$$f(z) = \sum_{n=-m}^{\infty} a_n q^n = \frac{a_{-m}}{q^m} + \frac{a_{-m+1}}{q^{m-1}} + \cdots + \frac{a_{-1}}{q} + a_0 + a_1 q + a_2 q^2 + \cdots$$

for some $m \in \mathbb{N}$.

(b) f is a *modular form of weight* k for $\mathrm{SL}(2,\mathbb{Z})$ if f is a modular function of weight k and it is *analytic everywhere on* $\mathbb{H}$ *and at the cusp* ∞ of $X(1)$. This means that $f(z)$ has a Taylor expansion

$$f(z) = \sum_{n=0}^{\infty} a_n q^n = a_0 + a_1 q + a_2 q^2 + \cdots .$$

Equivalently, f is analytic at the cusp ∞ if $|f(yi)|$ stays bounded as $y \to \infty$. The value of f at the cusp is equal to $\lim_{y\to\infty} f(yi)$.

(c) f is a *cusp form of weight* k for $\mathrm{SL}(2,\mathbb{Z})$ if f is a modular form of weight k and it *vanishes at the cusp* ∞ of $X(1)$. This means that the function $f(z)$ has a Taylor expansion

$$f(z) = \sum_{n=1}^{\infty} a_n q^n = a_1 q + a_2 q^2 + \cdots , \qquad \text{i.e., } a_0 = 0.$$

Equivalently, f is a cusp form if $\lim_{y\to\infty} f(yi) = 0$.

Remark 4.1.4. Exercise 4.5.5 shows that condition (2) in the definition of weakly modular function (resp. modular form below); i.e., Definition 4.1.3 above (resp. 4.2.1 below), is equivalent to saying that $f(z)(dz)^k$ is a differential k-form, invariant under the action of $\mathrm{SL}(2,\mathbb{Z})$ (resp. congruence subgroup Γ).

It is easy to show that there are no modular forms of odd weight for $\mathrm{SL}(2,\mathbb{Z})$ other than $f(z) = 0$; see Exercise 4.5.2. The Eisenstein series are our first and most important examples of modular forms of even weight $2k$ for $\mathrm{SL}(2,\mathbb{Z})$. Recall that we have already defined the Eisenstein series $G_{2k}(\Lambda)$ as functions of lattices (in Definition 3.2.3). Here we evaluate G_{2k} at the lattice $\Lambda_z = \langle z, 1\rangle$, with $z \in \mathbb{H}$. Therefore we may consider $G_{2k}(\Lambda_z) = G_{2k}(z)$ as a function of the complex variable z.

Proposition 4.1.5. *Let* $k \geq 2$ *and define a function of* $\tau \in \mathbb{H}$ *by*

$$G_{2k}(z) := G_{2k}(\langle z, 1\rangle) = \sum_{\substack{w\in\langle z,1\rangle \\ w\neq 0}} \frac{1}{w^{2k}} = \sum_{\substack{(m,n)\in\mathbb{Z}\times\mathbb{Z} \\ (m,n)\neq(0,0)}} \frac{1}{(mz+n)^{2k}}.$$

Then $G_{2k}(z)$ is a modular form of weight $2k$ for $\mathrm{SL}(2,\mathbb{Z})$, and the value of G_{2k} at the cusp ∞ of $X(1)$ is equal to $2\zeta(2k)$, where $\zeta(s)$ is the Riemann zeta function.

The proof is an exercise (Exercise 4.5.3 shows everything except the fact that G_{2k} is meromorphic on $\mathbb{H}$, which is an exercise in uniform convergence of series).

Definition 4.1.6. We say that a modular form is *normalized* if the first non-zero coefficient of its q-expansion is equal to 1.

The following proposition states the formula for the q-expansion of the Eisenstein series and its normalization.

Proposition 4.1.7. *Let $k \geq 2$, let $\zeta(s)$ be the Riemann zeta function, let $q = e^{2\pi i z}$ and let $\sigma_k(n) = \sum_{0<d|n} d^k$ be the sum of the k-th powers of positive divisors of n. Then, the q-expansion of the Eisenstein series $G_{2k}(z)$ is given by*

$$G_{2k}(z) = 2\zeta(2k) + \frac{2(2\pi i)^{2k}}{(2k-1)!} \sum_{n\geq 1} \sigma_{2k-1}(n) q^n.$$

Therefore, the normalized Eisenstein series is

$$E_{2k}(z) = \frac{1}{2\zeta(2k)} G_{2k}(z) = 1 - \frac{4k}{B_{2k}} \sum_{n\geq 1} \sigma_{2k-1}(n) q^n,$$

where B_{2k} is the $2k$-th Bernoulli number.

For a proof of this fact, see [**Kob93**], Ch. III, Prop. 6. The values of $\zeta(2k)$ can be computed in terms of Bernoulli numbers:

$$\zeta(2k) = \sum_{n=1}^{\infty} \frac{1}{n^{2k}} = -\frac{(2\pi i)^{2k} B_{2k}}{2(2k)!} \quad \text{for all } k \geq 1.$$

Next, we list a number of properties satisfied by modular forms.

Proposition 4.1.8. *Let $k, k' \geq 2$ be integers.*

(1) *Suppose that f and g are modular forms of weight k for $\mathrm{SL}(2,\mathbb{Z})$. Then, for all $\lambda, \mu \in \mathbb{C}$, the function $\lambda f(z) + \mu g(z)$ is also a modular form of weight k for $\mathrm{SL}(2,\mathbb{Z})$. Therefore, the set of all modular forms of weight k for $\mathrm{SL}(2,\mathbb{Z})$ is a vector space over $\mathbb{C}$.*

(2) *The set of all cusp forms of weight k for* $\mathrm{SL}(2,\mathbb{Z})$ *is a* $\mathbb{C}$*-linear subspace of the vector space of forms of weight k.*

(3) *Suppose that $f(z)$ and $g(z)$ are respectively modular forms of weight k and k' for* $\mathrm{SL}(2,\mathbb{Z})$*. Then the function $f(z)\cdot g(z)$ is a modular form of weight $k+k'$ for* $\mathrm{SL}(2,\mathbb{Z})$*.*

The proof is Exercise 4.5.4.

Definition 4.1.9. The $\mathbb{C}$-vector space of all modular forms of weight k for $\mathrm{SL}(2,\mathbb{Z})$ is denoted by $M_k(\mathrm{SL}(2,\mathbb{Z}))$. The subspace of cusp forms is denoted by $S_k(\mathrm{SL}(2,\mathbb{Z}))$.

Example 4.1.10. Let $g_2(z) = -15G_4(z)$, $g_3(z) = -35G_6(z)$, and define

$$\Delta(z) = -16\left(4(g_2(z))^3 + 27(g_3(z))^2\right).$$

Then Δ is a modular form of weight 12 for $\mathrm{SL}(2,\mathbb{Z})$. The modular form Δ is usually called the *modular discriminant*. $\Delta(z)$ has a simple zero at ∞ and no other zeros. The function

$$j(z) = 1728\frac{(4g_2(z))^3}{\Delta(z)}$$

is a modular function of weight 0 (as in Definition 4.1.1) but it is not a modular form. $j(z)$ is analytic on $\mathbb{H}$ but it is not analytic at ∞ because $\Delta(z)$ has a zero at ∞ but $g_2(z)$ does not vanish at ∞, so $j(z)$ has a pole. The function $j(z)$ is called the *modular j-invariant*. ■

Example 4.1.11. The modular forms $f_1(z) = E_{10}(z)$ and $f_2(z) = E_4(z)\cdot E_6(z)$ are both in the space $M_{10}(\mathrm{SL}(2,\mathbb{Z}))$. *A priori*, they are distinct modular forms. However, if we knew the dimension of $M_{10}(\mathrm{SL}(2,\mathbb{Z}))$ and the dimension was 1, then there should be a linear relationship between both f_1 and f_2. Thus, we need to know the dimensions of spaces of modular forms! ■

Theorem 4.1.12. *The dimension of $M_k(\mathrm{SL}(2,\mathbb{Z}))$ as a $\mathbb{C}$-vector space is finite and it is given by*

$$\dim_{\mathbb{C}}(M_k(\mathrm{SL}(2,\mathbb{Z}))) = \begin{cases} 0 & \text{if } k<0 \text{ or if } k \text{ is odd;} \\ \left[\frac{k}{12}\right] & \text{if } k\geq 0,\ k\equiv 2 \bmod 12; \\ \left[\frac{k}{12}\right]+1 & \text{otherwise.} \end{cases}$$

where $[x]$ is the greatest integer $\leq x$. Moreover, for all integers k, there is an isomorphism $\psi : M_{2k-12}(\mathrm{SL}(2,\mathbb{Z})) \cong S_{2k}(\mathrm{SL}(2,\mathbb{Z}))$ of $\mathbb{C}$-vector spaces given by $\psi(f(z)) = \Delta(z)f(z)$. In particular,

$$\dim_{\mathbb{C}}(S_{2k}(\mathrm{SL}(2,\mathbb{Z}))) = \dim_{\mathbb{C}}(M_{2k-12}(\mathrm{SL}(2,\mathbb{Z}))).$$

For a proof, see **[DS05]**, Theorem 3.5.2; **[Ser77]**, Ch. VII, Theorem 4; or **[Kob93]**, Ch. III, Proposition 9.

Example 4.1.13. Let $f_1(z) = E_{10}(z)$ and $f_2(z) = E_4(z) \cdot E_6(z)$, which are both in the space $M_{10}(\mathrm{SL}(2,\mathbb{Z}))$. By Theorem 4.1.12, the dimension of M_{10} as a $\mathbb{C}$-vector space is 1. Therefore, there exists $\lambda \in \mathbb{C}$ such that $f_1(z) = \lambda f_2(z)$. However, both $f_1(z)$ and $f_2(Z)$ are normalized modular forms, meaning that their first non-zero coefficient of their q-expansions equals 1. Hence, comparing their q-expansions, we conclude that $\lambda = 1$ and $E_{10}(z) = E_4(z)E_6(z)$.

The equality we just deduced, together with the q-expansion of the Eisenstein series (Prop. 4.1.7), can be rephrased as:

$$1 - \frac{20}{B_{10}} \sum_{n\geq 1} \sigma_9(n)q^n = (1 - \frac{8}{B_4} \sum_{k\geq 1} \sigma_3(k)q^k)(1 - \frac{12}{B_6} \sum_{h\geq 1} \sigma_5(h)q^h).$$

In particular, if we compare the coefficient of q^n on both sides, we obtain the following interesting conclusion:

$$\frac{20}{B_{10}}\sigma_9(n) = \frac{8}{B_4}\sigma_3(n) + \frac{12}{B_6}\sigma_5(n) - \frac{96}{B_4B_6} \sum_{j=1}^{n-1} \sigma_3(j)\sigma_5(n-j),$$

where $B_4 = -1/30$, $B_6 = 1/42$ and $B_{10} = 5/66$. We remind the reader that $\sigma_i(n) = \sum_{0<d|n} d^i$.

■

Example 4.1.14. Let $\Delta(z)$ be the modular discriminant form (which is a modular form of weight 12 for $\mathrm{SL}(2,\mathbb{Z})$), as in Example 4.1.10. It can be shown that

$$\Delta(z) = (2\pi)^{12} q \prod_{n=1}^{\infty} (1-q^n)^{24},$$

where $q = e^{2\pi iz}$ as usual. The Ramanujan τ-function is defined by the coefficients of the q-expansion of the normalized Δ function, i.e.,

$$(2\pi)^{-12}\Delta(z) = q\prod_{n=1}^{\infty}(1-q^n)^{24} = \sum_{n\geq 1}\tau(n)q^n.$$

Using the fact that $S_{12}(\mathrm{SL}(2,\mathbb{Z}))$ is 1-dimensional (by Theorem 4.1.12), one can show the following surprising congruence:

$$\tau(n) \equiv \sigma_{11}(n) \equiv \sum_{0<d|n} d^{11} \bmod 691 \quad \text{for all } n \geq 1.$$

A proof of this congruence is outlined in Exercise 4.5.7.

4.2. Modular forms for congruence subgroups

Definition 4.2.1. Let Γ be a congruence subgroup of $\mathrm{SL}(2,\mathbb{Z})$. A function $f : \mathbb{H} \to \mathbb{C} \cup \{\infty\}$ is *weakly modular of weight k* for Γ if:

(1) f is meromorphic on $\mathbb{H}$; and

(2) $f(Mz) = (cz+d)^k f(z)$ for any $M \in \Gamma$; i.e., $f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k f(z)$, where M is a matrix $(a,b;c,d) \in \Gamma$.

Since Γ is a congruence subgroup, there must be an $N \geq 1$ such that $\Gamma(N) \subseteq \Gamma$. If f is a weakly modular function of weight k for Γ, and $\Gamma(N) \subseteq \Gamma$, such that N is the smallest positive integer with this property, then we say that the *level of f is N*. If f is weakly modular of weight k and level N, then the matrix

$$T_N = \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix}$$

belongs to Γ. Therefore,

$$f(z) = f(T_N z) = f(z+N) \quad \text{for all} \quad z \in \mathbb{H}.$$

This periodicity of f means that, if we set $q_N = e^{2\pi iz/N}$, then we can express f as a Laurent series:

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (q_N)^n,$$

where a_n are the Fourier coefficients of f.

(a) We say that f is a *modular function of weight k* for Γ if f satisfies (1) and (2) above and f is *meromorphic at the cusps* of the modular curve $X(\Gamma) = \mathbb{H}^*/\Gamma$. This means that, for any matrix $M = (a, b; c, d) \in \mathrm{SL}(2, \mathbb{Z})$, the function $f_M(z) = (cz + d)^{-k} f\left(\frac{az+b}{cz+d}\right)$ has a Laurent expansion

$$f_M(z) = \sum_{n=-m}^{\infty} c_n (q_N)^n.$$

See Remark 4.2.2 below for a discussion about this requirement.

(b) f is a *modular form of weight k* for Γ if f is a modular function of weight k and it is *analytic on $\mathbb{H}$ and at the cusps* of $X(\Gamma)$. This means that, for all $M \in \mathrm{SL}(2, \mathbb{Z})$, the function $f_M(z)$ has a Taylor expansion

$$f_M(z) = \sum_{n=0}^{\infty} c_n (q_N)^n.$$

(c) f is a *cusp form of weight k* for Γ if f is a modular form of weight k and it *vanishes at all the cusps* of $X(\Gamma)$. This means that, for all $M \in \mathrm{SL}(2, \mathbb{Z})$, the function $f_M(z)$ has a Taylor expansion

$$f_M(z) = \sum_{n=1}^{\infty} c_n (q_N)^n, \quad \text{i.e., } c_0 = 0.$$

Remark 4.2.2. In part (a) we request that f should be meromorphic at the cusps of $X(\Gamma)$. It is simple to check whether f is meromorphic at the cusp ∞. Indeed, f is meromorphic at ∞ if it has a Laurent expansion of the form $f(z) = \sum_{n=-m}^{\infty} c_n (q_N)^n$. In order to check if f is meromorphic at another cusp, say $s = -\frac{d}{c} \in \mathbb{Q}$, we first do a change of variables that brings $-d/c$ to ∞. This is best accomplished by a linear fractional transformation $z \mapsto \frac{az+b}{cz+d}$ (see Prop. 3.4.3). The function $f_M(z)$ is precisely the result of this change of variables on $f(z)$. Thus, if $f_M(z)$ is meromorphic at ∞, then $f(z)$ is meromorphic at $-d/c$, so we just need to check the Laurent expansion at ∞ of $f_M(z)$.

Notice also that $X(\Gamma)$ only has a finite number of cusps, say $s_1, s_2, \ldots, s_n$, so one only needs to check the condition in part (a) of Defn. 4.2.1 for a finite number of matrices $M_1, M_2, \ldots, M_n$ such that M_i sends s_i to ∞.

Let Γ be a congruence subgroup. As in the case of modular forms for $\mathrm{SL}(2,\mathbb{Z})$ (cf. Proposition 4.1.8), the set of all modular forms of weight k for Γ is a $\mathbb{C}$-vector space and the subset of cusp forms is a linear subspace.

Proposition 4.2.3. *Let $k \geq 2$ and let Γ be a congruence subgroup. Let $M_k(\Gamma)$ be the set of all modular forms of weight k for Γ, and let $S_k(\Gamma)$ be the subset that consists of all cusp forms of weight k for Γ.*

(1) *$M_k(\Gamma)$ is a $\mathbb{C}$-vector space and $S_k(\Gamma)$ is a $\mathbb{C}$-linear subspace of $M_k(\Gamma)$.*

(2) *Let Γ' be another congruence subgroup contained in Γ, i.e., $\Gamma' \leq \Gamma$. Then any modular form $f(z)$ of weight k for Γ is also a modular form of the same weight for Γ'. Therefore, $M_k(\Gamma)$ is a $\mathbb{C}$-linear subspace of $M_k(\Gamma')$, and $S_k(\Gamma) \subseteq S_k(\Gamma')$.*

(3) *Let k and k' be positive integers. If $f(z) \in M_k(\Gamma)$ and $g(z) \in M_{k'}(\Gamma)$, then $(f \cdot g)(z)$ is a modular form in $M_{k+k'}(\Gamma)$.*

Remark 4.2.4. Let Γ be a congruence subgroup. Then $\Gamma \leq \mathrm{SL}(2,\mathbb{Z})$. Thus, Proposition 4.2.3 implies that $M_k(\mathrm{SL}(2,\mathbb{Z}))$ is always a $\mathbb{C}$-linear subspace of $M_k(\Gamma)$. Also, recall that $\Gamma(N) \leq \Gamma_1(N) \leq \Gamma_0(N)$. Therefore, $M_k(\Gamma_0(N)) \subseteq M_k(\Gamma_1(N)) \subseteq M_k(\Gamma(N))$.

Remark 4.2.5. Let $N \geq 1$ and let M be a positive divisor of N. Then

$$\Gamma_0(N) \leq \Gamma_0(M),\ \Gamma_1(N) \leq \Gamma_1(M),\ \text{ and }\ \Gamma(N) \leq \Gamma(M).$$

Therefore, for any $k \geq 1$, $M_k(\Gamma_0(M)) \subseteq M_k(\Gamma_0(N))$, $M_k(\Gamma_1(M)) \subseteq M_k(\Gamma_1(N))$, and $M_k(\Gamma(M)) \subseteq M_k(\Gamma(N))$.

Also, suppose that $N = MM'$, where $1 < M, M' < N$, so that M and M' are proper divisors of N. Suppose that $g(z) \in M_k(\Gamma(M))$. Then, it is an exercise (Exercise 4.5.9) to show that $f(z) := g(M'z)$ belongs to $M_k(\Gamma(N))$.

Definition 4.2.6. Let N, $k \geq 1$ be integers. A modular form $f(z)$ of weight k for $\Gamma(N)$ is said to be an *old form* if there is some positive divisor M of N such that $f(z)$ is a modular form in the space $M_k(\Gamma(M))$. The $\mathbb{C}$-linear subspace spanned by the set of all old forms of $M_k(\Gamma(N))$ is usually denoted by $M_k^{\mathrm{old}}(\Gamma(N))$. We also define

$$S_k^{\mathrm{old}}(\Gamma(N)) := M_k^{\mathrm{old}}(\Gamma(N)) \cap S_k(\Gamma(N)).$$

In Definition 4.3.2, we will define the space of *new cusp forms* as the orthogonal complement of $S_k^{\mathrm{old}}(\Gamma(N))$ in $S_k(\Gamma(N))$ with respect to the Petersson inner product (see Section 4.3).

As for $\mathrm{SL}(2,\mathbb{Z})$, the Eisenstein series (of level N) are the main non-trivial examples (compare the following definition with Proposition 4.1.5).

Definition 4.2.7. Let $k \geq 3$, $N \geq 1$ and let $a = (a_1, a_2) \in \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ be non-zero, i.e., $a \not\equiv (0,0) \bmod N$. We define the *Eisenstein series of level N and weight k*, corresponding to a, by

$$G_k^a(z) = \sum_{(m,n) \equiv a \bmod N} \frac{1}{(mz+n)^k},$$

where the sum is over all $(m,n) \in \mathbb{Z} \times \mathbb{Z}$ such that $m \equiv a_1$ and $n \equiv a_2 \bmod N$.

Notice that if $a = (0,0) \bmod N$, then we would have

$$\begin{aligned} G_k^a(z) &= \sum_{\substack{(m,n) \equiv (0,0) \bmod N \\ (m,n) \neq (0,0)}} \frac{1}{(mz+n)^k} \\ &= \sum_{\substack{(a,b) \in \mathbb{Z}^2 \\ (a,b) \neq (0,0)}} \frac{1}{(Naz+Nb)^k} = N^{-k} G_k(z), \end{aligned}$$

where $G_k(z)$ is the classical Eisenstein modular form of weight k for $\mathrm{SL}(2,\mathbb{Z})$.

Proposition 4.2.8. *Let $k \geq 3$, $N \geq 1$ and let $a = (a_1, a_2) \in \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ be non-zero. Then*

$$G_k^a(z) \in M_k(\Gamma(N)) \quad \textit{and} \quad G_k^{(0,a_2)}(z) \in M_k(\Gamma_1(N))$$

for any $a_2 \not\equiv 0 \bmod N$.

See Exercise 4.5.8 for a proof of the invariance under $\Gamma(N)$ and $\Gamma_1(N)$.

Remark 4.2.9. The Eisenstein series are very useful because most of the spaces we are discussing in this book have a basis formed by Eisenstein series, and we can calculate their q-expansions. For precise statements see **[Miy06]**, Chapter 7, or **[Ste07]**, Chapter 5 (in particular, see Section 5.3).

Remark 4.2.10. Let $k \geq 1$ and let Γ be a congruence subgroup. Then $M_k(\Gamma)$ is a *finite-dimensional* $\mathbb{C}$-vector space. The formulas for the dimension of the spaces of modular forms $M_k(\Gamma)$ and $S_k(\Gamma)$ can be found by calculating the genus and the number of cusps of the modular curve $X(\Gamma)$. Since we will not use these formulas here, we simply refer the reader to **[DS05]**, Theorem 3.5.1 and Figure 3.3 (page 108).

Example 4.2.11. Let $N = 11$ and let the weight be 2. The space $M_2(\Gamma_0(11))$ is a 2-dimensional $\mathbb{C}$-vector space with basis elements $\{f, g\}$ given by the q-expansions

$$\begin{aligned} f(q) &= q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 - 2q^{10} + O(q^{11}) \\ g(q) &= 1 + \frac{12}{5}q + \frac{36}{5}q^2 + \frac{48}{5}q^3 + \frac{84}{5}q^4 + \frac{72}{5}q^5 + \frac{144}{5}q^6 + O(q^7), \end{aligned}$$

where $q = e^{2\pi iz}$. Thus, we deduce that $S_2(\Gamma_0(11))$ is 1-dimensional, generated by $f(q)$.

Example 4.2.12. Let $N = 37$ and let the weight be 2. The space $M_2(\Gamma_0(37))$ is a 3-dimensional $\mathbb{C}$-vector space with basis elements $\{f, g, h\}$ given by the q-expansions:

$$\begin{aligned} f(q) &= q + q^3 - 2q^4 - q^7 - 2q^9 + 3q^{11} - 2q^{12} - 4q^{13} + O(q^{16}) \\ g(q) &= q^2 + 2q^3 - 2q^4 + q^5 - 3q^6 - 4q^9 - 2q^{10} + 4q^{11} + O(q^{12}) \\ h(q) &= 1 + \frac{2}{3}q + 2q^2 + \frac{8}{3}q^3 + \frac{14}{3}q^4 + 4q^5 + 8q^6 + \frac{16}{3}q^7 + O(q^8), \end{aligned}$$

where, once again, $q = e^{2\pi iz}$. Thus, we deduce that $S_2(\Gamma_0(37))$ is 2-dimensional, generated by $f(q)$ and $g(q)$.

4.3. The Petersson inner product

Let Γ and Γ' be congruence subgroups with $\Gamma' \subseteq \Gamma$. Let $\mathcal{F}(\Gamma') \subseteq \mathbb{C}$ be a fundamental domain for the modular curve $X(\Gamma') = \mathbb{H}^*/\Gamma'$, and suppose $f(z)$ and $g(z)$ are modular forms in $M_k(\Gamma')$ such that at least one of them is a cusp form in $S_k(\Gamma')$. We define the *Petersson inner product* of f and g by

$$\langle f, g\rangle := \frac{1}{[\overline{\Gamma} : \overline{\Gamma'}]} \int_{\mathcal{F}(\Gamma')} f(x+iy)\overline{g(x+iy)} y^k \frac{dxdy}{y^2} \in \mathbb{C},$$

where $\overline{\Gamma} = \Gamma/\{\pm \mathrm{Id}\}$ and $\overline{g(z)}$ is the complex conjugate of $g(z)$.

Remark 4.3.1. A number of remarks are in order:

(1) The Euclidean measure $dxdy$ on $\mathbb{C}$ has been replaced by the *hyperbolic measure* $\frac{dxdy}{y^2}$ on $\mathbb{H}$. This makes sense because the hyperbolic measure is *invariant* under $\mathrm{SL}(2,\mathbb{Z})$. This means that, if $M \in \mathrm{SL}(2,\mathbb{Z})$ and $\mathcal{F}$ is a region in $\mathbb{H}$, then

$$\int_{\mathcal{F}} \frac{dxdy}{y^2} = \int_{M\mathcal{F}} \frac{dxdy}{y^2},$$

where $M\mathcal{F} = \{Mz : z \in \mathcal{F}\}$ and the matrix $M = (a, b; c, d)$ acts on z as usual by $Mz = \frac{az+b}{cz+d}$.

(2) The integral in the definition of $\langle f(z), g(z)\rangle$ does not converge (in general) if neither f nor g is a cusp form.

(3) The Petersson inner product is a *Hermitian inner product* on $S_k(\Gamma')$, i.e.,

- $\langle f(z), g(z)\rangle$ linear in f:

$$\langle \lambda_1 f_1(z) + \lambda_2 f_2(z), g(z)\rangle = \lambda_1\langle f_1(z), g(z)\rangle + \lambda_2\langle f_2(z), g(z)\rangle$$

  for any $\lambda_1, \lambda_2 \in \mathbb{C}$ and any f_1, f_2 and $g \in S_k(\Gamma')$;
- $\langle f(z), g(z)\rangle$ is antilinear in g:

$$\langle f(z), \lambda_1 g_1(z) + \lambda_2 g_2(z)\rangle = \overline{\lambda_1}\langle f(z), g_1(z)\rangle + \overline{\lambda_2}\langle f(z), g_2(z)\rangle,$$

  where $\overline{\lambda}$ is the complex conjugate of $\lambda \in \mathbb{C}$;
- $\langle f(z), g(z)\rangle$ is conjugate-symmetric, i.e.,

$$\langle f(z), g(z)\rangle = \overline{\langle g(z), f(z)\rangle};$$

  and
- $\langle f(z), f(z)\rangle > 0$ for $f(z) \neq 0$.

Therefore, the Petersson inner product makes $S_k(\Gamma')$ into an inner product space.

(4) Suppose that Γ'' is another congruence subgroup with $\Gamma'' \leq \Gamma' \leq \Gamma$, and let $f(z)$ and $g(z)$ be modular forms for Γ'. Then, f and g may also be considered as modular forms for Γ'', and

$$\begin{aligned}\langle f, g\rangle &= \frac{1}{[\overline{\Gamma} : \overline{\Gamma'}]} \int_{\mathcal{F}(\Gamma')} f(x+iy)\overline{g(x+iy)} y^k \frac{dxdy}{y^2} \\ &= \frac{1}{[\overline{\Gamma} : \overline{\Gamma''}]} \int_{\mathcal{F}(\Gamma'')} f(x+iy)\overline{g(x+iy)} y^k \frac{dxdy}{y^2}.\end{aligned}$$

Thus, $\langle f, g\rangle$ gives the same value whether we consider f and g as modular forms for Γ' or for Γ''.

Definition 4.3.2. Let N, $k \geq 1$. Let $S_k^{\text{old}}(\Gamma(N))$ be the subspace of $S_k(\Gamma(N))$ of old forms, defined in Definition 4.2.6. We define a subspace of *new forms*, denoted by $S_k^{\text{new}}(\Gamma(N))$, as the orthogonal complement of $S_k^{\text{old}}(\Gamma(N))$ in $S_k(\Gamma(N))$ with respect to the Petersson inner product, i.e.,

$$\begin{aligned}&S_k^{\text{new}}(\Gamma(N)) := S_k^{\text{old}}(\Gamma(N))^{\perp} \\ &= \{f(z) \in S_k(\Gamma(N)) : \langle f(z), g(z)\rangle = 0 \text{ for all } g \in S_k^{\text{old}}(\Gamma(N))\}.\end{aligned}$$

Hence, $S_k(\Gamma(N))$ factors as:

$$\begin{aligned}S_k(\Gamma(N)) &= S_k^{\text{new}}(\Gamma(N)) \oplus S_k^{\text{old}}(\Gamma(N)) \\ &= S_k^{\text{old}}(\Gamma(N))^{\perp} \oplus S_k^{\text{old}}(\Gamma(N)).\end{aligned}$$

4.4. Hecke operators acting on cusp forms

Let $N \geq 1$ and let $S_k(\Gamma_0(N))$ be the space of cusp forms of weight k for $\Gamma_0(N)$. In this section we define a collection of $\mathbb{C}$-linear maps from $S_k(\Gamma_0(N))$ to $S_k(\Gamma_0(N))$ that will be very important in Chapter 5. In particular, we will be very interested in the eigenvalues and eigenvectors associated to these linear maps.

4.4.1. The w_N operator.

Definition 4.4.1. Let $N, k \geq 1$ and let $f(z)$ be a modular form in $S_k(\Gamma_0(N))$. The operator w_N on $S_k(\Gamma_0(N))$ is a linear map

$$w_N : S_k(\Gamma_0(N)) \to S_k(\Gamma_0(N))$$

defined by

$$(w_N(f))(z) = i^k \cdot \left(\sqrt{N}z\right)^{-k} \cdot f\left(\frac{-1}{Nz}\right).$$

Proposition 4.4.2. *Let $N, k \geq 1$ and let $f(z)$ be a modular form in $S_k(\Gamma_0(N))$. Then:*

- *$w_N(f)$ is also a modular form in $S_k(\Gamma_0(N))$;*
- *w_N is $\mathbb{C}$-linear, i.e., $w_N(\lambda \cdot f + \mu \cdot g) = \lambda \cdot w_N(f) + \mu \cdot w_N(g)$ for all $f, g \in S_k(\Gamma_0(N))$ and all $\lambda, \mu \in \mathbb{C}$; and*
- *The square of w_N is the identity, i.e., $w_N(w_N(f)) = f$.*

Therefore, $w_N^2 = \mathrm{Id}$ and the eigenvalues of w_N are $+1$ or -1. Thus, $S_k(\Gamma_0(N))$ can be expressed as the direct sum of the eigenspace corresponding to $+1$ plus the eigenspace corresponding to -1; i.e., if we define spaces

$$\begin{aligned} S_k^+(\Gamma_0(N)) &= \{f \in S_k(\Gamma_0(N)) : w_N(f) = f\}, \\ S_k^-(\Gamma_0(N)) &= \{f \in S_k(\Gamma_0(N)) : w_N(f) = -f\}, \end{aligned}$$

then $S_k(\Gamma_0(N))$ factors as

$$S_k(\Gamma_0(N)) = S_k^+(\Gamma_0(N)) \oplus S_k^-(\Gamma_0(N)).$$

We shall see in the next chapter that if $f(z) \in S_k(\Gamma_0(N))$ is in the $\varepsilon = \pm 1$ eigenspace, then the sign in the functional equation of the L-series attached to f is precisely ε. The proof of Proposition 4.4.2 is an exercise (Exercise 4.5.11).

4.4.2. The diamond operators. Let $\delta \in \mathbb{Z}$ and $N, k \geq 1$. The diamond operator $\langle \delta \rangle$ is a linear map from $M_k(\Gamma_1(N))$ to itself, defined as follows.

Definition 4.4.3. Let $\delta \in \mathbb{Z}$ be fixed. Let $M = (a, b; c, d)$ be a matrix in $\Gamma_0(N)$ such that $d \equiv \delta \bmod N$. The diamond operator $\langle\delta\rangle$ is a linear map $\langle\delta\rangle : M_k(\Gamma_1(N)) \to M_k(\Gamma_1(N))$ defined by

$$(\langle\delta\rangle f)(z) = (cz + d)^{-k} f(Mz) = (cz + d)^{-k} f\left(\frac{az + b}{cz + d}\right).$$

Exercise 4.5.12 shows that the definition of $\langle\delta\rangle$ does not depend on the choice of a matrix M. Thus, $\langle\delta\rangle$ is determined by the value of $\delta \bmod N$, so there are N distinct diamond operators, one for each value $0, 1, \ldots, N-1$. Notice that $\langle 1\rangle f = f$ is the identity operator, because we can pick $M = \mathrm{Id}$ in the definition of the diamond operator. Moreover, the following proposition shows that the diamond operators with $(\delta, N) = 1$ form a group under multiplication.

Proposition 4.4.4. *Let $N, k \geq 1$ be fixed and let $\delta, \delta' \in \mathbb{Z}$ with $(\delta\delta', N) = 1$. Then $\langle\delta'\rangle(\langle\delta\rangle f) = \langle\delta\rangle(\langle\delta'\rangle f) = \langle\delta'\delta\rangle f$. In particular, $\langle\delta\rangle^{\varphi(N)} = \langle 1\rangle = \mathrm{Id}$ and the eigenvalues of $\langle\delta\rangle$ must be roots of unity of order dividing $\varphi(N)$, where φ is the Euler phi function.*

The proof of this proposition is left to the reader: Exercise 4.5.13.

Let $\mu_{\varphi(N)}$ be the set of all roots of unity of order dividing $\varphi(N)$. Then, for each $\delta \in \mathbb{Z}$ and every $\zeta \in \mu_{\varphi(N)}$, there is an eigenspace of $M_k(\Gamma_1(N))$ formed by eigenvectors with eigenvalue ζ. More concretely, let $\delta \in \mathbb{Z}$ be fixed. Then, for each $\zeta \in \mu_{\varphi(N)}$, the set

$$M_k(\Gamma_1(N), \langle\delta\rangle, \zeta) = \{f(z) \in M_k(\Gamma_1(N)) : (\langle\delta\rangle f)(z) = \zeta \cdot f(z)\}$$

is a linear subspace of $M_k(\Gamma_1(N))$, which is the eigenspace for $\langle\delta\rangle$ formed by all eigenvectors with eigenvalue ζ. Furthermore, for each $\delta \in \mathbb{Z}$, the space of modular forms $M_k(\Gamma_1(N))$ can be decomposed as a direct sum of eigenspaces:

$$M_k(\Gamma_1(N)) = \bigoplus_{\zeta \in \mu_{\varphi(N)}} M_k(\Gamma_1(N), \langle\delta\rangle, \zeta).$$

As it turns out, one can show a much more interesting decomposition of $M_k(\Gamma_1(N))$, as follows.

Proposition 4.4.5. *Let $N, k \geq 1$ be fixed. For every group homomorphism $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ (i.e., a character) we define*

$$M_k(N, \chi) = \{f \in M_k(\Gamma_1(N)) : \langle\delta\rangle f = \chi(\delta) f \text{ for all } \delta \in (\mathbb{Z}/N\mathbb{Z})^\times\}.$$

Then $M_k(\Gamma_1(N)) = \bigoplus_\chi M_k(N,\chi)$*, where the direct sum is over all possible characters* χ *of* $(\mathbb{Z}/N\mathbb{Z})^\times$*.*

The reader should check (Exercise 4.5.14) that if χ_0 is the trivial character (i.e., $\chi_0(\delta) = 1$ for all $(\delta, N) = 1$), then $M_k(N,\chi_0) = M_k(\Gamma_0(N))$.

4.4.3. The T_n operators. Before we define the Hecke operators T_n, we need to define the auxiliary operators U_m and V_m.

Definition 4.4.6. Let $m \geq 1$ and let $f \in M_k(\Gamma_1(N))$. We define operators V_m and U_m by

$$(V_m(f))(z) = f(mz) \quad \text{and} \quad (U_m(f))(z) = \frac{1}{m}\sum_{j=0}^{m-1} f\left(\frac{z+j}{m}\right).$$

If f is given by a q-expansion $f(z) = \sum_{n\geq 0} a_n q^n$, then

$$V_m(f) = \sum_{n\geq 0} a_n q^{mn} \quad \text{and} \quad U_m(f) = \sum_{n \equiv 0 \bmod m} a_n q^{n/m}.$$

Recall that in Prop. 4.4.5 we defined spaces $M_k(N,\chi)$ by

$$M_k(N,\chi) = \{f \in M_k(\Gamma_1(N)) : \langle\delta\rangle f = \chi(\delta) f \text{ for all } \delta \in (\mathbb{Z}/N\mathbb{Z})^\times\}.$$

Definition 4.4.7. Let $f(z) \in M_k(N,\chi)$ and suppose $f(z)$ is given by a q-expansion $f(z) = \sum_{n\geq 0} a_n q^n$. Let $p \geq 2$ be a prime. We define an operator T_p by

$$T_p(f) = U_p(f) + \chi(p)p^{k-1}V_p(f),$$

where $\chi(p) = 0$ if $N \equiv 0 \bmod p$. Equivalently,

$$T_p(f(z)) = \sum_{n\geq 0} b_n q^n, \quad \text{such that} \quad b_n = a_{pn} + \chi(p)p^{k-1}a_{n/p}$$

and $a_{n/p} = 0$ if $n \not\equiv 0 \bmod p$. In particular, if χ_0 is trivial and $f \in M_k(N,\chi_0) = M_k(\Gamma_0(N))$, then

$$T_p(f) = U_p(f) + p^{k-1}V_p(f).$$

Next, we define Hecke operators T_n for all $n \geq 1$.

Definition 4.4.8. Let $f \in M_k(N,\chi)$. We define Hecke operators T_n for all $n \geq 1$ as follows:

- If $n = p \geq 2$ is a prime, then $T_p(f) = U_p(f) + \chi(p)p^{k-1}V_p(f)$ as before;
- If $n = p^r$ and $p|N$, then $T_{p^r} = (T_p)^r$, i.e., T_p composed r times with itself;
- If $n = p^r$ and $p \nmid N$, then T_{p^r} can be calculated using the following recurrence relation:

$$T_p \cdot T_{p^r} = T_{p^{r+1}} + p^{k-1}\langle p \rangle T_{p^{r-1}}.$$

- If $(n, m) = 1$, then $T_{nm}(f) = (T_n \cdot T_m)(f) = (T_m \cdot T_n)(f) = T_m(T_n(f))$.

Remark 4.4.9. There are several equivalent ways to define Hecke operators. T_n can be defined as above, or as a function on lattices, or as a double coset operator. See **[DS05]**, **[Kob93]** or **[Mil06]** for alternative definitions.

Every Hecke operator T_n defines a linear map $T_n : M_k(N, \chi) \to M_k(N, \chi)$. As in the case of the w_N and diamond operators, we are interested in the eigenvalues and eigenvectors of the operators T_n. Surprisingly, there exist eigenvectors f which satisfy $T_n(f) = \lambda_n f$ **for all** $n \geq 1$, i.e., f is an eigenvector **for all** Hecke operators simultaneously! These eigenvectors are of particular interest in the theory, as we shall see in the next chapter.

Definition 4.4.10. Let $f(z) \in M_k(N, \chi) \subset M_k(\Gamma_1(N))$. We say that $f(z)$ is an *eigenform* if f is an eigenvector **for all** Hecke operators T_n, $n \geq 1$, simultaneously. In other words, f is an eigenform if, for all $n \geq 1$, there exist eigenvalues $\lambda_n \in \mathbb{C}$ such that

$$T_n(f) = \lambda_n f.$$

Theorem 4.4.11 (Hecke; **[Kob93]**, Ch. III, Prop. 40)**.** *Let $k \geq 1$ and suppose that $f(z)$ is an eigenform in the space $M_k(N, \chi) \subset M_k(\Gamma_1(N))$, with $T_n(f) = \lambda_n f$ for all $n \geq 1$, for some $\lambda_n \in \mathbb{C}$. Suppose further that f has a q-expansion of the form $f(z) = \sum_{n \geq 0} a_n q^n$. Then:*

(1) *$a_1 \neq 0$ and $a_n = \lambda_n a_1$ for all $n \geq 1$; and*
(2) *if $a_0 \neq 0$, then the eigenvalues are given by the formula $\lambda_n = \sum_{d|n} \chi(d)d^{k-1}$.*

Example 4.4.12. Let $k \geq 2$ and let

$$E_{2k}(z) = \frac{1}{2\zeta(2k)} G_{2k}(z) = 1 - \frac{4k}{B_{2k}} \sum_{n \geq 1} \sigma_{2k-1}(n) q^n$$

be the (normalized) Eisenstein series of weight $2k$ for $\mathrm{SL}(2, \mathbb{Z})$, as in Proposition 4.1.7. We can write

$$\widehat{E}_{2k}(z) = -\frac{B_{2k}}{4k} E_{2k}(z) = -\frac{B_{2k}}{4k} + \sum_{n \geq 1} \sigma_{2k-1}(n) q^n.$$

Therefore, $a_1 = 1$ and $a_n = \sigma_{2k-1}(n) = \sum_{0<d|n} d^{2k-1}$. Since $\widehat{E}_{2k}$ is a modular form for $\mathrm{SL}(2, \mathbb{Z})$, it may also be considered as a form for $\Gamma_0(N)$ for any $N \geq 1$. Hence, $\widehat{E}_{2k} \in M_{2k}(N, \chi_0) = M_{2k}(\Gamma_0(N))$, where χ_0 is the trivial character of $(\mathbb{Z}/N\mathbb{Z})^\times$, and so

$$a_n = \sum_{0<d|n} \chi_0(d) d^{2k-1}$$

since $\chi_0(d) = 1$. Also notice that $a_0 = B_{2k}/4k \neq 0$, so $\widehat{E}_{2k}$ is not a cusp form. Hence, Hecke's theorem 4.4.11 suggests that E_{2k} may be an eigenform; that is, it suggests that E_{2k} is an eigenvector for all T_n, with $n \geq 1$, with eigenvalue a_n. In other words, $T_n(E_{2k}) = \sigma_{2k-1}(n) E_{2k}$ for all $n \geq 1$. This equality is left as an exercise for the reader (see Exercise 4.5.15). ■

Remark 4.4.13. One can show directly that the eigenvalues $\lambda_n = \sum_{d|n} \chi(d) d^{k-1}$ satisfy the recursive relation dictated by our definition of the Hecke operators (as in Definition 4.4.8). This is left as an exercise for the reader (Exercise 4.5.16).

Definition 4.4.14. In the notation of Theorem 4.4.11, an eigenform $f \in M_k(N, \chi) \subset M_k(\Gamma_1(N))$ is said to be *normalized* if $a_1 = 1$ (c.f. Definition 4.1.6).

Remark 4.4.15. The reader may have noticed that we now have two different types of normalizations (see Definition 4.1.6). Depending on the application, it is convenient to normalize modular forms in one way or another. For instance, in Example 4.1.13 we have normalized $a_0 = 1$ for convenience, while Theorem 4.4.11 shows that when working with eigenforms it is very useful to normalize them so that $a_1 = 1$.

Example 4.4.16. It follows from Hecke's theorem that, if $k \geq 2$, there is a unique normalized eigenform of $M_{2k}(\Gamma_0(N))$ that is not a cusp form, and it is precisely the Eisenstein series $\widehat{E}_{2k}$, by Example 4.4.12 (and Exercise 4.5.15). ■

It is a crucial fact in the theory that one can find a basis for $S_k^{\mathrm{new}}(\Gamma_1(N))$ such that every element of the basis is an eigenform. Recall that we defined spaces of new forms and old forms (Definitions 4.2.6 and 4.3.2) such that

$$S_k(\Gamma(N)) = S_k^{\mathrm{new}}(\Gamma(N)) \oplus S_k^{\mathrm{old}}(\Gamma(N)) = S_k^{\mathrm{old}}(\Gamma(N))^{\perp} \oplus S_k^{\mathrm{old}}(\Gamma(N))$$

where new and old forms are orthogonal with respect to the Petersson inner product. We define $S_k^{\mathrm{new}}(\Gamma_1(N)) = S_k(\Gamma_1(N)) \cap S_k^{\mathrm{new}}(\Gamma(N))$, and define similarly the old space for $\Gamma_1(N)$.

Theorem 4.4.17 (Atkin, Lehner **[AL70]**, Li **[Li75]**; see also **[DS05]**, §5.8)**.** *The spaces of modular forms $S_k^{new}(\Gamma_1(N))$ and $S_k^{old}(\Gamma_1(N))$ are stable under w_N, the diamond operators and T_n for all $n \geq 1$. Furthermore, the space $S_k^{new}(\Gamma_1(N))$ has an orthonormal basis that consists of new normalized eigenforms for the Hecke operators w_N and T_n for all $n \geq 1$. In other words, the new forms of weight k for $\Gamma_1(N)$ have a basis $\{f_1, \ldots, f_d\}$ such that*

$$\langle f_i, f_j \rangle = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j \end{cases} \qquad w_N(f_i) = \pm f_i \quad \text{and} \quad T_n f_i = \lambda_{n,i} f_i,$$

where $\langle \cdot, \cdot \rangle$ is the Petersson inner product for all $n \geq 1$ and all $1 \leq i \leq d$, for some eigenvalues $\lambda_{n,i} \in \mathbb{C}$.

Definition 4.4.18. A normalized eigenform $f \in S_k^{\mathrm{new}}(\Gamma_1(N))$, which is an eigenvector for all Hecke, w_N and diamond operators simultaneously, is called a *newform* (not to be confused with simply a new form).

We remind the reader that in the previous theorem and definition the word "normalized" means that $a_1 = 1$, as in Definition 4.4.14.

4.5. Exercises

Exercise 4.5.1. The goal of this problem is to show that the modularity condition (2) in Definition 4.1.3 works "as one would hope" under matrix multiplication. Let $M = (a, b; c, d)$ and $M' = (a', b'; c', d')$ be matrices in a congruence subgroup $\Gamma \leq \mathrm{SL}(2, \mathbb{Z})$, and let $MM' = (a'', b''; c'', d'')$ be their product. Let $k \geq 1$ and let $f(z)$ be a function. Show that:

(1) $(MM')z = M(M'z)$, i.e., $\frac{a''z+b''}{c''z+d''} = \frac{a'(Mz)+b'}{c'(Mz)+d'}$, where $Mz = \frac{az+b}{cz+d}$;

(2)
$$(cz+d)^{-k}\left((c'(Mz)+d')^{-k} f\left(\frac{a'(Mz)+b'}{c'(Mz)+d'}\right)\right)$$
$$= (c''z+d'')^{-k} f\left(\frac{a''z+b''}{c''z+d'}\right).$$

Exercise 4.5.2. Let $f(z)$ be a weakly modular function of weight k for $\mathrm{SL}(2, \mathbb{Z})$ with k odd. Show that $f(z) = 0$ for all $z \in \mathbb{H}$. (Hint: show that $f(z) = -f(z)$ for all $z \in \mathbb{H}$.)

Exercise 4.5.3. Let $G_{2k}(z)$ be the Eisenstein series
$$G_{2k}(z) := \sum_{(0,0)\neq(m,n)\in\mathbb{Z}\times\mathbb{Z}} \frac{1}{(mz+n)^{2k}}.$$

(1) Show that $G_{2k}\left(\frac{az+b}{cz+d}\right) = (cz+d)^{2k} G_{2k}(z)$ for all $(a, b; c, d) \in \mathrm{SL}(2, \mathbb{Z})$.

(2) Show that
$$\lim_{y\to\infty} \sum_{(0,0)\neq(m,n)\in\mathbb{Z}\times\mathbb{Z}} \frac{1}{(myi+n)^{2k}} = 2\zeta(2k)$$
where $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ is the Riemann zeta function. (Hint: you may assume that the convergence is uniform [can you prove this?], and so the limit can be brought inside the summation.)

Exercise 4.5.4. Prove Propositions 4.1.8 and 4.2.3.

Exercise 4.5.5. Let f be a modular form of weight k for a congruence subgroup $\Gamma \subseteq \mathrm{SL}(2,\mathbb{Z})$. Show that:

$$f(z) = f\left(\frac{az+b}{cz+d}\right)\left(\frac{\partial}{\partial z}\left(\frac{az+b}{cz+d}\right)\right)^k \text{ for any } M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma.$$

(Note: this exercise shows that condition (2) in the definition of modular form, Definition 4.2.1, is equivalent to saying that $f(z)(dz)^k$ is a differential k-form, invariant under the action of Γ.)

Exercise 4.5.6. Show that the modular discriminant $\Delta(z)$, as defined in Example 4.1.10, is a cusp form of weight 12 for $\mathrm{SL}(2,\mathbb{Z})$.

Exercise 4.5.7. Let $\Delta(z)$ be the modular discriminant form, and let $E_{2k}(z)$ be the normalized Eisenstein series of weight $2k$ for $\mathrm{SL}(2,\mathbb{Z})$.

(1) Show that $M = M_{12}(\mathrm{SL}(2,\mathbb{Z}))$ is a 2-dimensional vector space, and $\{\Delta(z)\}$ is a basis of the cusp forms $S_{12}(\mathrm{SL}(2,\mathbb{Z}))$.

(2) Show that E_{12} and E_6^2 belong to M, and

$$E_{12} - E_6^2 = \lambda\Delta, \quad \text{where } \lambda = \frac{(2\pi)^{-12}\cdot 2^6\cdot 3^5\cdot 7^2}{691}.$$

(3) Use the q-expansions of E_6, E_{12} and Δ to write an expression for $\tau(n)$ in terms of $\sigma_5(n)$ and $\sigma_{11}(n)$.

(4) Show that $\tau(n) \equiv \sigma_{11}(n) \equiv \sum_{0<d|n} d^{11} \bmod 691$ for all $n \geq 1$.

Exercise 4.5.8. Let $k \geq 3$, $N \geq 1$ and let $a = (a_1, a_2) \in \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ be non-zero. Let $G_k^a(z)$ be the Eisenstein series of Definition 4.2.7, and let $M = (a, b; c, d) \in \mathrm{SL}(2,\mathbb{Z})$.

(1) Show that

$$(cz+d)^k G_k^a\left(\frac{az+b}{cz+d}\right) = G_k^{aM}(z),$$

where $aM = (a_1a + a_2c,\ a_1b + a_2d) \bmod \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$.

(2) Show that if $M \in \Gamma(N)$, then $a \equiv aM \bmod N$, and if $M \in \Gamma_1(N)$, then $(0, a_2) \equiv (0, a_2)M \bmod N$.

(3) Conclude that $G_k^a(z)$ is modular for $\Gamma(N)$ and the function $G_k^{(0,a_2)}(z)$ is modular for $\Gamma_1(N)$.

Exercise 4.5.9. Let $N \geq 1$ and suppose that $N = M \cdot M'$, where $1 < M, M' < N$ so that M and M' are proper divisors of N. Suppose that $g(z) \in M_k(\Gamma(M))$. Show that $f(z) := g(M'z) \in M_k(\Gamma(N))$. Also, show that the same conclusion holds if we replace $\Gamma(N)$ by $\Gamma_1(N)$. Hint:

$$\begin{pmatrix} aM' & bM' \\ cN & d \end{pmatrix} = \begin{pmatrix} a & bM' \\ cM & d \end{pmatrix} \cdot \begin{pmatrix} M' & 0 \\ 0 & 1 \end{pmatrix}.$$

Exercise 4.5.10. Show that the Petersson inner product is a Hermitian inner product (see Remark 4.3.1 for the properties of a Hermitian product).

Exercise 4.5.11. This exercise proves the properties of w_N that are claimed in Proposition 4.4.2. Let $N, k \geq 1$.

(1) Verify the following identity of matrices:

$$\begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -c/N \\ -Nb & a \end{pmatrix} \cdot \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}.$$

Conclude that $\Gamma_0(N)$ is preserved under conjugation by the matrix $(0, -1; N, 0)$.

(2) Use the matrix identity in (1) to show that, if the matrix $M = (a, b; c, d)$ is in $\Gamma_0(N)$ and $z' = Mz = \frac{az+b}{cz+d}$, then $(w_N(f))(z') = (cz + d)^k \cdot (w_N(f))(z)$. Thus, $w_N(f) \in S_k(\Gamma_0(N))$.

(3) Show that $w_N(w_N(f)) = f$ for all $f \in S_k(\Gamma_0(N))$. Hence $w_N(w_N) = \mathrm{Id}$, the eigenvalues of w_N are all ± 1, and $S_k(\Gamma_0(N))$ factors into the direct sum of eigenspaces

$$S_k(\Gamma_0(N)) = S_k^+(\Gamma_0(N)) \oplus S_k^-(\Gamma_0(N)).$$

Exercise 4.5.12. Let $N \geq 1$, $\delta \in \mathbb{Z}$ and let $M = (a, b; c, d) \in \Gamma_0(N)$ with $\delta \equiv d \bmod N$.

(1) Show that, for any $f \in M_k(\Gamma_1(N))$, the modular form $\langle \delta \rangle f$ is also modular of weight k under the action of $\Gamma_1(N)$ (where $\langle \delta \rangle$ is as in Definition 4.4.3).

(2) Let $M' = (a', b'; c', d') \in \Gamma_0(N)$ be another matrix with $\delta \equiv d \equiv d' \bmod N$, and let $f \in M_k(\Gamma_1(N))$. Show that

$$(cz + d)^{-k} f(Mz) = (c'z + d')^{-k} f(M'z)$$

for any $z \in \mathbb{H}$. (Hint: show that $M'M^{-1} \in \Gamma_1(N)$.)

Exercise 4.5.13. Prove Proposition 4.4.4. (Hint: use Exercise 4.5.1.)

Exercise 4.5.14. Let $M_k(N, \chi)$ be as in Proposition 4.4.5. Show that if χ_0 is the trivial character (i.e., $\chi_0(\delta) = 1$ for all $\gcd(\delta, N) = 1$), then $M_k(N, \chi_0) = M_k(\Gamma_0(N))$.

Exercise 4.5.15. Let $k \geq 2$ and let

$$\widehat{E}_{2k}(z) = -\frac{B_{2k}}{4k} + \sum_{n \geq 1} \sigma_{2k-1}(n) q^n$$

be the (renormalized) Eisenstein series of weight $2k$ for $\mathrm{SL}(2, \mathbb{Z}) = \Gamma_0(1)$, as in Example 4.4.12. Show that:

(1) $T_p(\widehat{E}_{2k}) = \sigma_{2k-1}(p)\widehat{E}_{2k}$ for all primes $p \geq 2$. (Hint: here $N = 1$ and $\chi = \chi_0$ is trivial.)

(2) $T_{p^r}(\widehat{E}_{2k}) = \sigma_{2k-1}(p^r)\widehat{E}_{2k}$ for any prime $p \geq 2$ and $r \geq 1$. (Hint: use induction and the recursive definition of T_{p^r}.)

(3) $T_n(\widehat{E}_{2k}) = \sigma_{2k-1}(n)\widehat{E}_{2k}$ for all $n \geq 1$. (Hint: σ_{2k-1} is multiplicative for relatively prime integers, i.e., $\sigma_{2k-1}(mn) = \sigma_{2k-1}(m)\sigma_{2k-1}(n)$ for $(m, n) = 1$.)

Exercise 4.5.16. Let $N > 1$ be a natural number, and let $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ be a character. For each $k \geq 2$, we define

$$\sigma_n = \sum_{d|n} \chi(d) d^{k-1}.$$

Show that

(1) If $(n, m) = 1$, then $\sigma_n \sigma_m = \sigma_{nm}$.

(2) If p is a prime with $\gcd(N, p) = 1$, then

$$\sigma_p \sigma_{p^r} = \sigma_{p^{r+1}} + \chi(p) p^{k-1} \sigma_{p^{r-1}}.$$

Chapter 5

L-functions

In this chapter we define the L-functions attached to elliptic curves and modular forms, and we investigate when an elliptic curve and a modular form could have the same L-function.

5.1. The L-function of an elliptic curve

Let E be an elliptic curve over $\mathbb{Q}$ given by a minimal model (as in Definition 2.6.3):

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with coefficients $a_i \in \mathbb{Z}$. For p a prime in $\mathbb{Z}$ of good reduction for $E/\mathbb{Q}$, we define N_p as the number of points in the reduction of the curve modulo p, i.e., the number of points in $E(\mathbb{F}_p)$. In other words, N_p is the number of points in

$$\{O\} \cup \{(x,y) \in {\mathbb{F}_p}^2 \colon y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6 \equiv 0 \bmod p\}$$

where O is the point at infinity (see Section 2.6 and, in particular, Hasse's theorem 2.6.11). Also, let $a_p = p + 1 - N_p$. We define the

local factor at p of the L-series to be

$$L_p(T) = \begin{cases} 1 - a_pT + pT^2, \text{ if } E \text{ has good reduction at } p, \\ 1 - T, \text{ if } E \text{ has split multiplicative reduction at } p, \\ 1 + T, \text{ if } E \text{ has non-split multiplicative reduction at } p, \\ 1, \text{ if } E \text{ has additive reduction at } p. \end{cases}$$

Definition 5.1.1. The L-function of the elliptic curve E is defined to be

$$L(E, s) = \prod_{p \geq 2} \frac{1}{L_p(p^{-s})},$$

where the product is over all primes $p \geq 2$ and $L_p(T)$ is the local factor defined above. $L(E, s)$ is sometimes called the Hasse-Weil L-function of $E/\mathbb{Q}$.

Remark 5.1.2. The product that defines $L(E, s)$ converges and gives an analytic function for all $\Re(s) > 3/2$. This follows from Hasse's bound (Theorem 2.6.11), which implies that $|a_p| \leq 2\sqrt{p}$. However, far more is true. Indeed, mathematicians conjectured that $L(E, s)$ should have an analytic continuation to the whole complex plane and that it must satisfy a functional equation relating the values of $L(E, s)$ and $L(E, 2-s)$. For the precise functional equation see Theorem 5.1.9 below.

Example 5.1.3. Let $E/\mathbb{Q}$ be the elliptic curve with equation

$$y^2 + y = x^3 - x^2 - 10x - 20.$$

This is a minimal model for $E/\mathbb{Q}$, and its discriminant is $\Delta_E = -11^5$. Therefore, $p = 11$ is the only prime of bad reduction for $E/\mathbb{Q}$, and the reduction is split multiplicative (see the discussion about E_3 in Example 2.6.7). Therefore,

$$L(E, s) = \left(\frac{1}{1 - 11^{-s}}\right) \cdot \prod_{\substack{p \geq 2 \\ p \neq 11}} \frac{1}{1 - a_p p^{-s} + p^{1-2s}}.$$

When expanded, the L-series attached to E has the form

$$L(E, s) = 1 - \frac{2}{2^s} - \frac{1}{3^s} + \frac{2}{4^s} + \frac{1}{5^s} + \frac{2}{6^s} - \frac{2}{7^s} - \frac{2}{9^s} - \frac{2}{10^s} + \frac{1}{11^s} + \cdots.$$

In general, one can always write $L(E, s) = \sum_{n \geq 1} a_n n^{-s}$, where the a_n are characterized in Proposition 5.1.5 below. ■

Example 5.1.4. Let $E/\mathbb{Q} : y^2 = x^3 - 11x^2 + 385$. The curve E has bad additive reduction at 2 and 11, split multiplicative at 5 and non-split multiplicative at 7 and 461. Thus, by definition

$$\begin{aligned} L(E,s) &= \big((1-5^{-s})(1+7^{-s})(1+461^{-s})\big)^{-1} \\ &\quad \cdot \prod_{\substack{\text{primes } p \\ p\neq 2,5,7,11,461}} \frac{1}{1-a_p p^{-s}+p^{1-2s}} \\ &= 1 - \frac{2}{3^s} + \frac{1}{5^s} - \frac{1}{7^s} + \frac{1}{9^s} + \frac{2}{13^s} - \frac{2}{15^s} - \frac{5}{17^s} + \frac{2}{21^s} \cdots . \end{aligned}$$

Proposition 5.1.5. *Let $E/\mathbb{Q}$ be an elliptic curve, and let $L(E,s)$ be its L-function. Define Fourier coefficients a_n for all $n \geq 1$ as follows. Let $a_1 = 1$. If $p \geq 2$ is prime, we define*

$$a_p = \begin{cases} p+1-N_p & \text{if } E \text{ has good reduction at } p; \\ 1 & \text{if } E \text{ has split multiplicative reduction at } p; \\ -1 & \text{if } E \text{ has non-split multiplicative reduction at } p; \\ 0 & \text{if } E \text{ has additive reduction at } p. \end{cases}$$

If $n = p^r$ for some $r \geq 1$, we define a_{p^r} recursively using the relation

$$a_p \cdot a_{p^r} = a_{p^{r+1}} + p \cdot a_{p^{r-1}} \quad \text{if } E/\mathbb{Q} \text{ has good reduction at } p$$

and $a_{p^r} = (a_p)^r$ if $E/\mathbb{Q}$ has bad reduction at p. Finally, if $(m,n) = 1$, then we define $a_{mn} = a_m \cdot a_n$. Then the L-function of E can be written as the series

$$L(E,s) = \sum_{n\geq 1} \frac{a_n}{n^s}.$$

The proof is left as an exercise (Exercise 5.7.2).

Remark 5.1.6. Notice that the recurrence formula $a_p \cdot a_{p^r} = a_{p^{r+1}} + p \cdot a_{p^{r-1}}$ (and $a_{p^r} = (a_p)^r$ in the bad reduction case) is strikingly similar to the recurrence relation defining the Hecke operators T_{p^r} for $k = 2$, and also the recurrence relation satisfied by the eigenvalues of an eigenform (see Definition 4.4.8, Remark 4.4.13 and Exercise 4.5.16). This is one of the first pieces of evidence that the L-function of an elliptic curve may be connected to a modular form.

Before we write down the functional equation for $E/\mathbb{Q}$, we need one more ingredient: the conductor of $E/\mathbb{Q}$. For each prime $p \in \mathbb{Z}$, we define the quantity f_p as follows:

$$f_p = \begin{cases} 0, \text{ if } E \text{ has good reduction at } p, \\ 1, \text{ if } E \text{ has multiplicative reduction at } p, \\ 2, \text{ if } E \text{ has additive reduction at } p, \text{ and } p \neq 2, 3, \\ 2 + \delta_p, \text{ if } E \text{ has additive reduction at } p = 2 \textit{ or } 3, \end{cases}$$

where δ_p is a technical invariant (see **[Sil94]**, Ch. IV, §10; the invariant δ_p describes whether there is wild ramification in the action of the inertia group at p of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the Tate module $T_p(E)$).

Definition 5.1.7. The *conductor* $N_{E/\mathbb{Q}}$ of $E/\mathbb{Q}$ is defined to be

$$N_{E/\mathbb{Q}} = \prod_p p^{f_p},$$

where the product is over all primes and the exponents f_p are defined as above.

Example 5.1.8. Let us see some examples of conductors.

(1) Let $E/\mathbb{Q}$: $y^2 + y = x^3 - x^2 + 2x - 2$. The primes of bad reduction for E are $p = 5$ and 7. The reduction at $p = 5$ is additive, while the reduction at $p = 7$ is multiplicative. Hence $N_{E/\mathbb{Q}} = 25 \cdot 7 = 175$.

(2) As we saw above, the curve $y^2 + y = x^3 - x^2 - 10x - 20$ has split multiplicative reduction at $p = 11$ and the reduction is good elsewhere. Thus, the conductor is 11.

(3) The curves $E_A : y^2 + y = x^3 - x$ and $E_B : y^2 + y = x^3 + x^2 - 23x - 50$ are two non-isomorphic curves with conductor equal to 37.

Theorem 5.1.9 (Functional equation)**.** *The L-series $L(E, s)$ has an analytic continuation to the entire complex plane, and it satisfies the following functional equation. Define*

$$\Lambda(E, s) = (N_{E/\mathbb{Q}})^{s/2}(2\pi)^{-s}\Gamma(s)L(E, s),$$

where $N_{E/\mathbb{Q}}$ is the conductor of E and $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt$ is the Gamma function. Then

$$\Lambda(E,s) = w \cdot \Lambda(E, 2-s) \quad \textit{with } w = \pm 1.$$

The number $w = w(E/\mathbb{Q})$ in the functional equation is usually called the *root number* of E, and it has an important conjectural meaning (see the next section on the Birch and Swinnerton-Dyer conjecture). Theorem 5.1.9 was proved in 1999, since it follows from the Taniyama-Shimura-Weil conjecture 5.4.5, which was proved by work of Wiles, Taylor-Wiles, and Breuil, Conrad, Diamond and Taylor.

5.2. The Birch and Swinnerton-Dyer conjecture

Figure 1. Bryan Birch (left) and Sir Peter Swinnerton-Dyer (right). Photograph courtesy of William Stein.

Conjecture 5.2.1 (Birch and Swinnerton-Dyer)**.** *Let E be an elliptic curve over $\mathbb{Q}$, and let $L(E,s)$ be the L-function attached to E. Then:*

(1) *$L(E,s)$ has a zero at $s=1$ of order equal to the rank R_E of $E(\mathbb{Q})$. In other words, the Taylor expansion of $L(E,s)$ at $s=1$ is of the form*

$$L(E,s) = C_0 \cdot (s-1)^{R_E} + C_1 \cdot (s-1)^{R_E+1} + C_3 \cdot (s-1)^{R_E+2} + \cdots$$

where C_0 is a non-zero constant.

(2) *The residue of $L(E,s)$ at $s=1$, i.e., the coefficient C_0, has a concrete expression in terms of invariants of $E/\mathbb{Q}$. More concretely,*

$$C_0 = \lim_{s\to 1} \frac{L(E,s)}{(s-1)^{R_E}} = \frac{|Ш| \cdot \Omega_E \cdot \mathrm{Reg}(E/\mathbb{Q}) \cdot \prod_p c_p}{|E_{\mathrm{torsion}}(\mathbb{Q})|^2}.$$

The invariants that appear in the conjectural formula for the residue are listed below:

- R_E is the (free) rank of $E(\mathbb{Q})$ (see Section 2.7).
- $\Omega_E = \int_{E(\mathbb{R})} \left|\frac{dx}{y}\right|$ is either the real period or twice the real period of a minimal model for E, depending on whether $E(\mathbb{R})$ is connected.
- $|Ш|$ is the order of the Shafarevich-Tate group of $E/\mathbb{Q}$ (we defined the 2-torsion of Sha, $Ш_2$, in Section 2.11).
- $\mathrm{Reg}(E/\mathbb{Q})$ is the elliptic regulator of $E(\mathbb{Q})$, as in Definition 2.8.4.
- $|E(\mathbb{Q})_{\mathrm{torsion}}|$ is the number of torsion points on $E/\mathbb{Q}$, including the point at infinity $\mathcal{O}$ (see Section 2.5).
- c_p is an elementary local factor, equal to the cardinality of $E(\mathbb{Q}_p)/E_0(\mathbb{Q}_p)$, where $E_0(\mathbb{Q}_p)$ is the set of points in $E(\mathbb{Q}_p)$ whose reduction modulo p is non-singular in $E(\mathbb{F}_p)$. Notice that if p is a prime of good reduction for $E/\mathbb{Q}$, then $c_p = 1$, so $c_p \neq 1$ only for finitely many primes p. The number c_p is called the *Tamagawa number* of E at p.

In 1974 ([**Tat74**], p. 198), John Tate wrote about the BSD conjecture: "*This remarkable conjecture relates the behavior of a function L at a point where it is not at present known to be defined ($s=1$) to the order of a group* (Ш) *which is not known to be finite!*" Tate is referring to the fact that, when the conjecture was first proposed, the analytic continuation of $L(E,s)$ was not known, and we did not know whether Ш was ever finite (nowadays we know many examples where Ш is finite, but it is still not known for all elliptic curves).

Example 5.2.2. Let $E/\mathbb{Q}$ be an elliptic curve. By Theorem 5.1.9, the function $L(E,s) = \sum_{n\geq 1} a_n n^{-s}$ has an analytic continuation to $\mathbb{C}$. In particular, if we restrict our attention to real values t, then $L(E,t)$

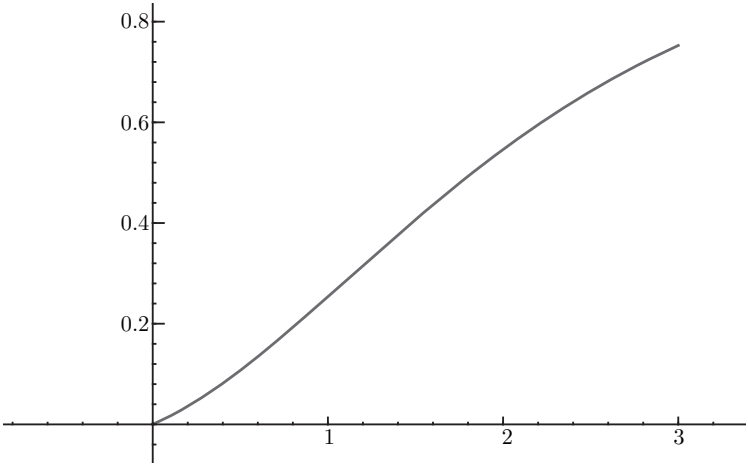


Figure 2. $L(E_0, t)$ for $E_0 : y^2 + y = x^3 - x^2 - 10x - 20$ and $-1 \leq t \leq 3$.

is a real-valued function. Since $L(E, s)$ is analytic, $L(E, t)$ should be continuous and (infinitely) differentiable. Let E_r, for $r = 0, 1, 2$ and 3, be elliptic curves defined by

$$\begin{aligned} E_0 &: y^2 + y = x^3 - x^2 - 10x - 20, \quad E_1 : y^2 + y = x^3 - x \\ E_2 &: y^2 + y = x^3 + x^2 - 2x, \quad E_3 : y^2 + y = x^3 - 7x + 6. \end{aligned}$$

The reader can check that the rank of E_r is precisely r. In Figures 2 through 5 we show the graphs of $L(E_r, t)$ for $-1 \leq t \leq 3$. Notice that the function $L(E_r, t)$ seems to have a zero of order r at $t = 1$, in agreement with the BSD conjecture. ■

Example 5.2.3. Let $E/\mathbb{Q} : y^2 = x^3 - 1156x$. Recall that in Examples 2.10.4 and 2.11.2 we calculated $R_E = 2$, $E(\mathbb{Q})_{\text{torsion}} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ and $Ш_2 = \{(1, 1)\}$ (here $Ш_2$ is just the 2-torsion of Ш). A non-trivial calculation yields $Ш = Ш_2 = \{(1, 1)\}$. Figure 6 provides the values of all the invariants that appear in the BSD conjecture. Thus,

$$\frac{|Ш| \cdot \Omega_E \cdot \text{Reg}(E/\mathbb{Q}) \cdot \prod_p c_p}{|E(\mathbb{Q})_{\text{torsion}}|^2} = 6.3851519548\ldots.$$

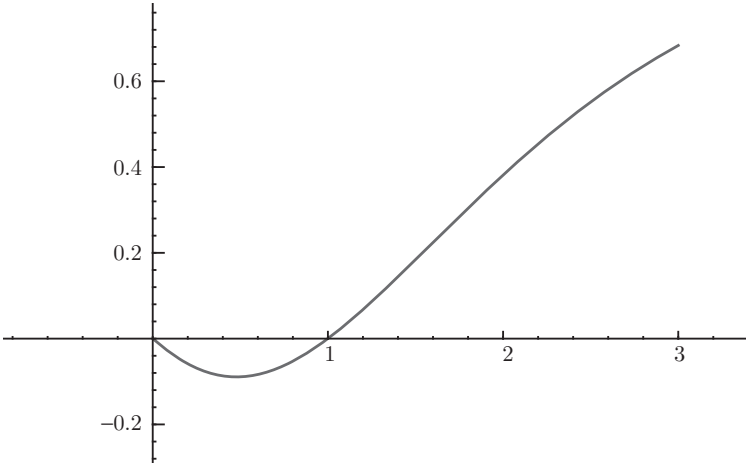


Figure 3. $L(E_1, t)$ for $E_1 : y^2 + y = x^3 - x$ and $-1 \leq t \leq 3$.

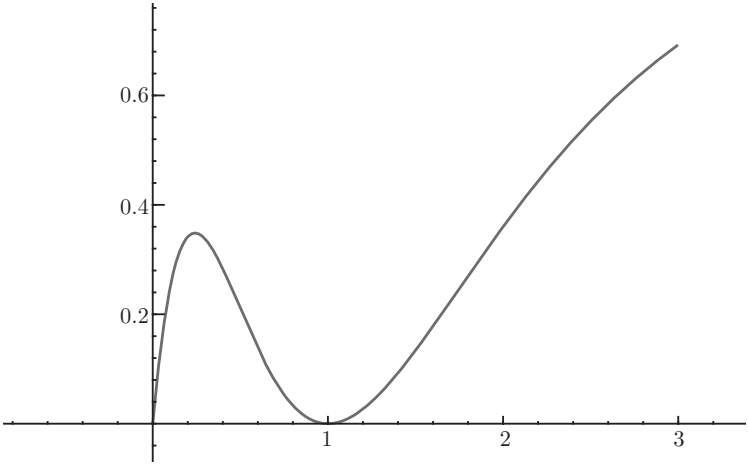


Figure 4. $L(E_2, t)$ for $E_2 : y^2 + y = x^3 + x^2 - 2x$ and $-1 \leq t \leq 3$.

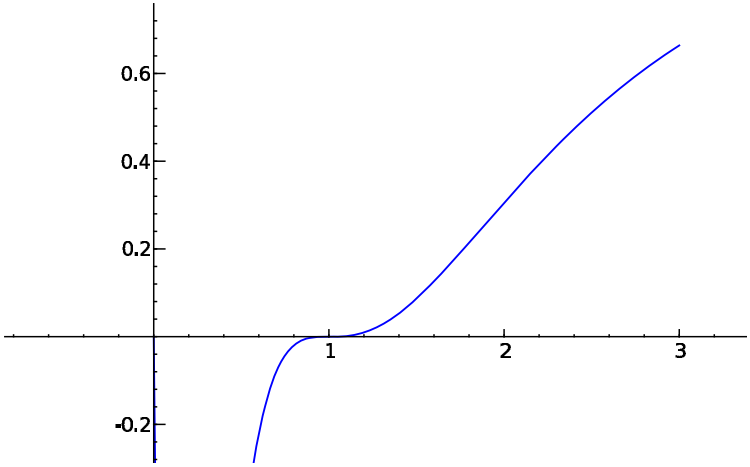


Figure 5. $L(E_3, t)$ for $E_3 : y^2 + y = x^3 - 7x + 6$ and $-1 \leq t \leq 3$.

$E/\mathbb{Q} : y^2 = x^3 - 1156x$	
R_E	2, $\langle P = (-16, 120), Q = (-2, 48)\rangle$
$\lvert Ш\rvert$	1
Ω_E	$0.8993583214\ldots$
$\mathrm{Reg}(E/\mathbb{Q})$	$\det \mathcal{H}(\{P, Q\}) = 7.0996751824\ldots$
$E(\mathbb{Q})_{\mathrm{torsion}}$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \cong \langle (0,0), (34,0)\rangle$
$\prod_{p \geq 2} c_p$	$c_2 \cdot c_{17} = 4 \cdot 4$

Figure 6. BSD data for the curve $E/\mathbb{Q} : y^2 = x^3 - 1156x$.

We can also calculate the value $L(E, 1)$ and the values of the derivatives $L'(E, 1)$ and $L''(E, 1)$; i.e., we can approximate numerically these values. For instance, one can use Sage (see Appendix A.3). For a technical description of the algorithms involved, see **[Dok04]**. Once we have calculated these values, we can write the first few terms

$E/\mathbb{Q} : y^2 = x^3 - 6724x$	
R_E	0
$\lvert Ш\rvert$	4
Ω_E	$0.5791156343\ldots$
$\mathrm{Reg}(E/\mathbb{Q})$	$\det \mathcal{H}(\{\}) = 1$
$E(\mathbb{Q})_{\mathrm{torsion}}$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \cong \langle (0,0), (81,0) \rangle$
$\prod_{p \geq 2} c_p$	$c_2 \cdot c_{41} = 4 \cdot 4$

Figure 7. BSD data for the curve $E/\mathbb{Q} : y^2 = x^3 - 6724x$.

of the Taylor expansion of $L(E,s)$ around $s = 1$.

$$\begin{aligned} L(E,s) &\approx 9.508 \cdot 10^{-24} - (2.374 \cdot 10^{-23}) \cdot (s-1) \\ &\quad + (6.3851519548) \cdot (s-1)^2 + \cdots . \end{aligned}$$

Therefore, our approximate calculation suggests that $L(E,s)$ has a zero of order 2 at $s = 1$ and the residue is $6.3851519548\ldots$, in perfect agreement with the BSD conjecture (at least up to the given precision). ■

Example 5.2.4. Let $E/\mathbb{Q} : y^2 = x^3 - 6724x$. Recall that Examples 2.10.5 and 2.11.3 *suggest* that $R_E = 0$, $E(\mathbb{Q})_{\mathrm{torsion}} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ and $|Ш_2| = 4$. A non-trivial calculation reveals that R_E is indeed 0 and $|Ш| = |Ш_2| = 4$. Figure 7 provides the values of all the invariants that appear in the BSD conjecture. Thus,

$$\frac{|Ш| \cdot \Omega_E \cdot \mathrm{Reg}(E/\mathbb{Q}) \cdot \prod_p c_p}{|E(\mathbb{Q})_{\mathrm{torsion}}|^2} = 2.3164625374\ldots .$$

We can approximate the first few terms of the Taylor expansion of $L(E,s)$ around $s = 1$.

$$\begin{aligned} L(E,s) &\approx 2.3164625374 - (7.8248271660) \cdot (s-1) \\ &\quad + (25.7352635691) \cdot (s-1)^2 + \cdots . \end{aligned}$$

Therefore, our approximate calculation suggests that $L(E,s)$ does not vanish at $s = 1$, and $L(E,1) = 2.3164625374\ldots$, again in perfect agreement with the BSD conjecture. ■

The following is an easy consequence of the BSD conjecture (Exercise 5.7.3). Recall that the root number of E is the sign in the functional equation of $L(E, s)$.

Conjecture 5.2.5 (Parity Conjecture)**.** *The root number of E, denoted by $w = w(E/\mathbb{Q})$, indicates the parity of the rank of the elliptic curve; i.e., $w = 1$ if and only if the rank R_E is even, and $w = -1$ iff the rank is odd. Equivalently,*

$$w = (-1)^{\mathrm{ord}_{s=1} L(E,s)} = (-1)^{\mathrm{rank}(E(\mathbb{Q}))}$$

or $\mathrm{ord}_{s=1} L(E, s) \equiv \mathrm{rank}(E(\mathbb{Q})) \bmod 2$.

See Exercise 5.7.3.

Definition 5.2.6. Let $E/\mathbb{Q}$ be an elliptic curve, and let $L(E, s)$ be the L-function attached to E. The *analytic rank* of $E(\mathbb{Q})$ is defined to be the order of vanishing of $L(E, s)$ at $s = 1$, i.e.,

$$\mathrm{rank}_{\mathrm{an}}(E/\mathbb{Q}) := \mathrm{ord}_{s=1} L(E, s).$$

In other words, $\mathrm{rank}_{\mathrm{an}}(E/\mathbb{Q})$ is the order of the zero of $L(E, s)$ at $s = 1$.

Thus, the first part of the BSD conjecture is the statement that the analytic rank equals the (algebraic) free rank of the Mordell-Weil group $E(\mathbb{Q})$.

Example 5.2.7. Let $E/\mathbb{Q} : y^2 = x^3 - 157^2 x$. Recall that Proposition 1.1.3 says that the rational points on $E/\mathbb{Q}$ with $y \neq 0$ give right triangles of area 157, so if we find a single non-trivial point on E we prove that $n = 157$ is a congruent number (as defined in Example 1.1.2).

Comparing values of $\Lambda(s)$ and $\Lambda(2 - s)$, we calculate the root number $w = w(E/\mathbb{Q}) = -1$. Thus, the parity conjecture suggests that $E(\mathbb{Q})$ has odd rank, therefore ≥ 1, and so $E(\mathbb{Q})$ must be infinite. However, a computer search only yields the trivial 2-torsion points $(0, 0)$, $(157, 0)$ and $(-157, 0)$. We can calculate values of $L(E, s)$ and its derivatives at $s = 1$ and write down an approximate Taylor expansion:

$$L(E, s) \approx (11.4259445007) \cdot (s - 1) - (49.9773214816) \cdot (s - 1)^2 + \cdots .$$

Hence, the BSD conjecture suggests that $R_E = 1$ and

$$\frac{|Ш| \cdot \Omega_E \cdot \mathrm{Reg}(E/\mathbb{Q}) \cdot \prod_p c_p}{|E(\mathbb{Q})_{\mathrm{torsion}}|^2} = 11.4259445007\ldots. \tag{5.1}$$

If we believe that $R_E = 1$ and we write P for a generator of $E(\mathbb{Q})$ modulo torsion, then one can show that $Ш_2$ must be trivial (and, in fact, $Ш$ is trivial as well, but this is much tougher to prove). Some other invariants are easy to calculate:

$$\Omega_E = 0.4185259488\ldots, \ \prod_{p\geq 2} c_p = c_2 \cdot c_{157} = 2 \cdot 4, \ |E(\mathbb{Q})_{\mathrm{torsion}}| = 4.$$

However, $\mathrm{Reg}(E/\mathbb{Q}) = \langle P, P\rangle = 2 \cdot \widehat{h}(P)$ is difficult to calculate because we do not know P (here $\widehat{h}$ is the canonical height). But we can solve for $\mathrm{Reg}(E/\mathbb{Q})$ in Eq. (5.1) and obtain

$$\mathrm{Reg}(E/\mathbb{Q}) = 2 \cdot \widehat{h}(P) = 54.6008892938\ldots$$

and $\widehat{h}(P) = 27.3004446469\ldots$. That's a huge height! Recall that

$$\widehat{h}(P) \approx \frac{1}{2} \log \max\{\mathrm{num}(x(P)), \mathrm{den}(x(P))\}$$

and so $\max\{|\mathrm{num}(x(P))|, |\mathrm{denom}(x(P))|\} \approx e^{54.6} \approx 5.157 \cdot 10^{23}$. This calculation gives us a rough idea of the size of the numerator and denominator of the x coordinate. With the help of homogeneous spaces, and looking for points in the correct height range, we can succeed at finding P. Its coordinates $P = (x(P), y(P))$ are:

$$x(P) = -\frac{16613623166818526754080 4}{28256306942511458580 25},$$

$$y(P) = \frac{167661624456834335404812111469782006}{150201095200135518108761470235125}$$

and the canonical height of P is precisely $27.3004446469\ldots$, as predicted by the Birch and Swinnerton-Dyer conjecture. ■

There has been a great amount of research on the BSD conjecture, but the progress in the general case over $\mathbb{Q}$ is minimal (a lot is known about BSD for elliptic curves over function fields). The conjecture has been verified for many elliptic curves (for instance, see [**GJPST09**], [**Mil10**]), but there is little evidence in the form of proven theorems. The following result is the strongest piece of evidence proved to date.

Theorem 5.2.8 (Gross-Zagier, Kolyvagin)**.** *Let $E/\mathbb{Q}$ be an elliptic curve of algebraic rank R_E. Suppose that the analytic rank of $E/\mathbb{Q}$ is ≤ 1, i.e., $\mathrm{ord}_{s=1} L(E,s) \leq 1$. Then:*

(1) *The first part of BSD holds for $E/\mathbb{Q}$, i.e.,*

$$R_E = \mathrm{rank}(E(\mathbb{Q})) = \mathrm{rank}_{an}(E/\mathbb{Q}) = \mathrm{ord}_{s=1} L(E,s).$$

(2) *The Shafarevich-Tate group Ш associated to $E/\mathbb{Q}$ is finite.*

5.3. The L-function of a modular (cusp) form

Let $N, k \geq 1$ and let $f(z)$ be a cusp form of weight $2k$ for the congruence subgroup $\Gamma_0(N)$, i.e., $f(z) \in S_{2k}(\Gamma_0(N))$ in the notation of Section 4.2 (and, in particular, Prop. 4.2.3). For any $N \geq 1$, the matrix $T = (1,1;0,1)$ belongs to $\Gamma_0(N)$, and therefore $f(z) = f(z+1)$ for all $z \in \mathbb{H}$. Moreover, $f(z)$ is a cusp form and so f vanishes at all the cusps of $\mathbb{H}^*/\Gamma_0(N)$, and in particular it vanishes at ∞. Hence $f(z)$ has a q-expansion expression of the form

$$f(z) = \sum_{n \geq 1} a_n q^n,$$

where $q = e^{2\pi i z}$ for some coefficients $a_n \in \mathbb{C}$.

Definition 5.3.1. The L-function attached to a cusp form $f(z) = \sum_{n\geq 1} a_n q^n \in S_k(\Gamma_0(N))$ is defined by

$$L(f,s) = \sum_{n \geq 1} a_n n^{-s} = \sum_{n \geq 1} \frac{a_n}{n^s} = a_1 + \frac{a_2}{2^s} + \frac{a_3}{3^s} + \cdots.$$

Example 5.3.2. Let $N = 11$ and $k = 1$. The space $M_2(\Gamma_0(11))$ is a 2-dimensional $\mathbb{C}$-vector space with basis elements $\{f, g\}$ given in Example 4.2.11. In particular, $S_2(\Gamma_0(11))$ is generated by

$$f(q) \quad = \quad q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 - 2q^{10} + O(q^{11}),$$

where $q = e^{2\pi i z}$. Hence the L-function associated to f is

$$L(f,s) = 1 - \frac{2}{2^s} - \frac{1}{3^s} + \frac{2}{4^s} + \frac{1}{5^s} + \frac{2}{6^s} - \frac{2}{7^s} - \frac{2}{9^s} - \frac{2}{10^s} + \cdots.$$

The very attentive reader might recognize these few terms as the first few terms in the L-function $L(E,s)$ that appeared in Example 5.1.3, where $E/\mathbb{Q}$ is the elliptic curve with equation $y^2 + y = x^3 - x^2 -$

$10x - 20$. Are they truly the same L-series? Further calculations show that all terms agree as we increase the precision. We will see that the Taniyama-Shimura-Weil conjecture 5.4.5, i.e., the modularity theorem, implies that $L(E, s) = L(f, s)$. Notice that the conductor of $E/\mathbb{Q}$ is precisely $N = 11$, as we saw in Example 5.1.8. ■

Example 5.3.3. Let $N = 37$ and $k = 1$. In Example 4.2.12 we described the space $S_2(\Gamma_0(37))$ with basis elements $\{f, g\}$ given by the q-expansions

$$\begin{aligned} f(q) &= q + q^3 - 2q^4 - q^7 - 2q^9 + 3q^{11} - 2q^{12} - 4q^{13} + O(q^{16}), \\ g(q) &= q^2 + 2q^3 - 2q^4 + q^5 - 3q^6 - 4q^9 - 2q^{10} + 4q^{11} + O(q^{12}). \end{aligned}$$

The L-functions attached to f and g are

$$\begin{aligned} L(f, s) &= 1 + \frac{1}{3^s} - \frac{2}{4^s} - \frac{1}{7^s} - \frac{2}{9^s} + \frac{3}{11^s} - \frac{2}{12^s} - \frac{4}{13^s} + \dots, \\ L(g, s) &= \frac{1}{2^s} + \frac{2}{3^s} - \frac{2}{4^s} + \frac{1}{5^s} - \frac{3}{6^s} - \frac{4}{9^s} - \frac{2}{10^s} + \frac{4}{11^s} + \dots. \end{aligned}$$

Now, let E_A and E_B be the elliptic curves of conductor 37 described in Example 5.1.8. Then

$$L(E_B, s) = 1 + \frac{1}{3^s} - \frac{2}{4^s} - \frac{1}{7^s} - \frac{2}{9^s} + \frac{3}{11^s} - \frac{2}{12^s} - \frac{4}{13^s} + \dots$$

and, indeed, we shall see that $L(f, s) = L(E_B, s)$. How about E_A?

$$L(E_A, s) = 1 - \frac{2}{2^s} - \frac{3}{3^s} + \frac{2}{4^s} - \frac{2}{5^s} + \frac{6}{6^s} - \frac{1}{7^s} + \frac{6}{9^s} + \frac{4}{10^s} - \frac{5}{11^s} + \dots$$

so $L(E_A, s) \neq L(g, s)$ or $L(f, s)$. Is there some form $F(z) \in S_k(\Gamma_0(37))$ such that $L(E_A, s) = L(F, s)$? If so, $F(q)$ must be a linear combination $\lambda \cdot f(q) + \mu \cdot g(q)$ for some $\lambda, \mu \in \mathbb{C}$. After a quick look at the first few coefficients of the q-expansions of f and g, and those of the series $L(E_A, s)$, one can check that, if some F works, then it must be $F(q) = f(q) - 2g(q)$, and indeed

$$(f - 2g)(q) = 1 - 2q^2 - 3q^3 + 2q^4 - 2q^5 + 6q^6 - q^7 + 6q^9 + 4q^{10} - 5q^{11} + O(q^{12})$$

and so

$$L(f - 2g, s) = 1 - \frac{2}{2^s} - \frac{3}{3^s} + \frac{2}{4^s} - \frac{2}{5^s} + \frac{6}{6^s} - \frac{1}{7^s} + \frac{6}{9^s} + \frac{4}{10^s} - \frac{5}{11^s} + \dots$$

Once again, we shall see that the Taniyama-Shimura-Weil conjecture implies the equality $L(f - 2g, s) = L(E_A, s)$. ■

5.4. The Taniyama-Shimura-Weil conjecture

In Examples 5.3.2 and 5.3.3, we have seen examples of elliptic curves $E/\mathbb{Q}$ of conductor N and modular forms $f \in S_2(\Gamma(N))$ such that the L-functions $L(E,s)$ and $L(f,s)$ seem to be identical.

Definition 5.4.1. We say that an elliptic curve $E/\mathbb{Q}$ is *modular* if there is a cusp form $f(z)$ such that

$$L(E,s) = L(f,s).$$

In the second half of the 20th century, many mathematicians grew increasingly interested in the question of whether every elliptic curve over $\mathbb{Q}$ is modular. However, early on, it was noticed that not every cusp form comes from an elliptic curve.

Notice that if E is modular and $L(E,s) = L(f,s) = \sum_{n\geq 1} a_n n^{-s}$, then a_p must equal $p+1-N_p$ when p is a prime of good reduction for E and, in general, a_n must coincide with those values defined in Proposition 5.1.5. Hence, for a given elliptic curve, there is a clear candidate for a cusp form f associated to the elliptic curve E.

Definition 5.4.2. Let $E/\mathbb{Q}$ be an elliptic curve. We define the *potential cusp form* associated to E to be a function $f_E : \mathbb{H} \to \mathbb{C}$ defined by its q-expansion

$$f_E(q) = \sum_{n\geq 1} a_n q^n,$$

where $q = e^{2\pi i z}$ and the a_n are defined in Proposition 5.1.5 (for instance, if $E/\mathbb{Q}$ has good reduction at p, then $a_p = p+1-N_p$).

It is *very far from clear* that f_E is a modular form. Let us suppose for a moment that f_E is indeed a modular form and $L(E,s) = L(f_E,s)$. What kind of modular form should f_E be?

(1) The examples suggest that, first of all, f_E must be a cusp form of weight 2 for $\Gamma_0(N)$, where $N = N_E$ is the conductor of $E/\mathbb{Q}$;

(2) If $L(E,s) = L(f_E,s)$, then, by the functional equation for $L(E,s)$ in Theorem 5.1.9, the L-function associated to f_E, that is $L(f_E,s)$, must also satisfy a functional equation;

(3) If $L(E,s) = L(f_E,s)$, then $L(f_E,s)$ must have an Euler product, since $L(E,s)$ has one. We say that $L(s) = \sum_{n\geq 1} a_n n^{-s}$ has an Euler product if it can be written as a product $L(s) = \prod_{p\geq 2} L_p(s)$ over all primes $p \geq 2$. Clearly, $L(E,s)$ is defined as an Euler product, so $L(f_E,s)$ must have an Euler product as well.

The work of Hecke characterizes which cusp forms in $S_2(\Gamma_0(N))$ satisfy a functional equation and which cusp forms have an Euler product. Recall that in Proposition 4.4.2 we defined ± 1-spaces of S_2 such that

$$S_2(\Gamma_0(N)) = S_2^+(\Gamma_0(N)) \oplus S_2^-(\Gamma_0(N)).$$

Theorem 5.4.3 (Hecke; **[DS05]**, §5.10)**.** *Let $N, k \geq 1$ and $f(z) \in S_{2k}(\Gamma_0(N))$ be a cusp form such that $f(z)$ is an eigenvector for the operator w_N, i.e., $f(z) \in S_{2k}^{\varepsilon}(\Gamma_0(N))$ for $\varepsilon = +1$ or -1. Then $L(f,s)$ has an analytic continuation to $\mathbb{C}$. Moreover, if we define*

$$\Lambda(f,s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(f,s),$$

where $\Gamma(s)$ is the Gamma function, then $\Lambda(f,s)$ satisfies the functional equation

$$\Lambda(f,s) = \varepsilon \cdot \Lambda(f, 2-s).$$

Recall (Definition 4.4.10) that we say that $f(z) = \sum_{n\geq 0} a_n q^n$ is an eigenform if f is an eigenvector **for all** Hecke operators T_n, $n \geq 1$, simultaneously. We say that $f(z)$ is a normalized eigenform if $a_1 = 1$.

Theorem 5.4.4 (Hecke; **[DS05]**, §5.9)**.** *Let $N, k \geq 1$. Let $f(z)$ be a normalized eigenform of weight $2k$ for $\Gamma_0(N)$ such that $T_p(f) = \lambda_p \cdot f$ for every prime $p \geq 2$. Then $L(f,s)$ has an Euler product of the form*

$$L(f,s) = \prod_{p|N} \frac{1}{1-\lambda_p p^{-s}} \prod_{p\nmid N} \frac{1}{1-\lambda_p p^{-s} + p^{2k-1-2s}}.$$

Now we may use Hecke's theorems to narrow down which cusp forms may be associated to elliptic curves. Suppose that $E/\mathbb{Q}$ is an elliptic curve with conductor N and let us assume that the potential cusp form f_E associated to E is indeed a cusp form. Then f_E must verify the following properties:

(1) $f_E(z) \in S_2(\Gamma_0(N))$. The level of $f_E(z)$ should be precisely N and not lower; otherwise f would correspond to a curve of lower conductor. Thus, we require $f_E(z) \in S_2^{\text{new}}(\Gamma_0(N))$. Note that the functional equation of f_E determines N, the conductor/level.

(2) $f_E(z)$ must be in one of the ε-spaces of cusp forms, i.e.,

$$f_E \in S_2^{\text{new}}(\Gamma_0(N)) \cap S_2^{\varepsilon}(\Gamma_0(N))$$

for $\epsilon = +1$ or -1.

(3) $f_E(z)$ must be a normalized eigenform in $S_2^{\text{new}}(\Gamma_0(N))$, and it needs to be an eigenvector for w_N as well. Therefore, $f_E(z)$ is a normalized newform (Definition 4.4.18).

Taniyama, Shimura and Weil are credited with the following formulation of the modularity conjecture.

Conjecture 5.4.5 (Taniyama-Shimura-Weil)**.** *A series of the form* $L(s) = \sum_{n\geq 1} a_n n^{-s}$ *with* $a_n \in \mathbb{Z}$ *is the L-function* $L(E, s)$ *of an elliptic curve* $E/\mathbb{Q}$ *of conductor* N *if and only if* $L(s) = L(f, s)$ *is the L-function of a normalized newform of weight* 2 *for* $\Gamma_0(N)$*.*

The conjecture of Taniyama, Shimura and Weil was proved in several stages.

- Eichler and Shimura ([**Shi73**], Ch. 7, Thm. 7.14) showed one of the directions of the equivalence in the conjecture: if $f(z)$ is a normalized newform of weight 2 for $\Gamma_0(N)$, then there exists an elliptic curve $E_f/\mathbb{Q}$ such that $L(f, s) = L(E_f, s)$.
- Wiles [**Wil95**] and Taylor and Wiles [**TW95**] proved the Taniyama-Shimura-Weil conjecture when $E/\mathbb{Q}$ is *semistable* (i.e., if the conductor N_E is square-free or, equivalently, when $E/\mathbb{Q}$ does not have any primes of bad additive reduction). This was the case that was needed to finalize the proof of Fermat's last theorem (see Section 5.5).
- Finally, Breuil, Conrad, Diamond and Taylor [**BCDT01**] showed that the conjecture is true for all elliptic curves over $\mathbb{Q}$.

The Taniyama-Shimura-Weil conjecture is nowadays frequently called the modularity theorem. We conclude this section with an important equivalent formulation of the TSW conjecture:

Theorem 5.4.6 (Modularity theorem)**.** *Let $E/\mathbb{Q}$ be an elliptic curve of conductor N, and let $X_0(N)$ be given by an algebraic model over $\mathbb{Q}$ (see Remark 3.6.4). Then there is a surjective algebraic map of curves $\Psi_{E,N} : X_0(N) \to E$ defined over $\mathbb{Q}$. (The map $\Psi_{E,N}$ is called a modular parametrization of E.)*

5.5. Fermat's last theorem

Theorem 5.5.1. *The equation $x^n + y^n = z^n$ has no solutions in integers x, y, z with $xyz \neq 0$, whenever $n > 2$.*

Figure 8. Andrew J. Wiles (right) and his Ph.D. advisor, John H. Coates (left).

Suppose that n, u, v and w are integers such that $n > 2$, $uvw \neq 0$ and

$$u^n + v^n = w^n.$$

Therefore, either n is divisible by 4, i.e., $n = 4k$ with $k \geq 1$, and $(u^k)^4 + (v^k)^4 = (w^k)^4$, or there is a prime divisor $p \geq 3$ of n, with

$n = ph$ and $h \geq 1$, such that $(u^h)^p + (v^h)^p = (w^h)^p$. Fermat showed that the equation $x^4 + y^4 = z^4$ has no solutions $x, y, z \in \mathbb{Z}$ with $xyz \neq 0$, so we conclude that $x^p + y^p = z^p$ must have an integer solution for some prime $p \geq 3$ and $xyz \neq 0$.

Thus, let us suppose that $p \geq 3$ and $a^p + b^p = c^p$, with $a, b, c \in \mathbb{Z}$ and $abc \neq 0$. However, we know that this is not possible for $p = 3, 5$ or 7.

- Leonhard Euler is generally credited for the proof of the $p = 3$ case (although his solution, in 1770, had a major gap). Kausler (1802), Legendre (1823) and many others have also published proofs of this case.
- The case of $p = 5$ was first shown (independently) by Legendre and Dirichlet, around 1825.
- The proof of Fermat's last theorem for $p = 7$ is due to Lamé, published in 1839.

Hence, we may assume that $p \geq 11$. It is worth pointing out that, in 1846, Ernst Kummer proved Fermat's last theorem for *regular* primes. Not all primes are regular: we know that there are infinitely many irregular primes (the first few irregular primes are 37, 59, 67, 101, 103, 131, 149, ...), but it is widely believed that there are also infinitely many regular primes. In 1984, the proof of Mordell's conjecture (now known as Faltings' theorem; see the paragraph on *Higher degree* in Section 2.1) was announced which shows that, for a fixed $n > 2$, $x^n + y^n = z^n$ may have at most a finite number of relatively prime integer solutions.

The strategy that led to the first (correct) proof of Fermat's last theorem was layed out by Frey [**Fre86**] and Serre [**Ser87**]. Let $p \geq 11$ and suppose a, b, c are relatively prime integers with $a^p + b^p = c^p$ and $abc \neq 0$. In 1984, Frey discovered that the elliptic curve

$$E : y^2 = x(x - a^p)(x + b^p)$$

would be semistable with conductor $N_E = \prod_{\ell | abc} \ell$ (see Exercise 5.7.5) and would satisfy some other technical properties. Moreover, Frey claimed that such a curve $E/\mathbb{Q}$ could not be modular; i.e., there is no weight 2 normalized newform $f \in S_2(\Gamma_0(N_E))$ such that

$L(f,s) = L(E,s)$. The problem with the modularity of E was made precise by Serre, and Ribet [**Rib90**] proved in 1986 that, indeed, E cannot be modular.

Finally, in 1995, Wiles [**Wil95**] and Taylor and Wiles [**TW95**] proved the Taniyama-Shimura-Weil conjecture 5.4.5 for all semistable elliptic curves $E/\mathbb{Q}$. Therefore, $E : y^2 = x(x-a^p)(x+b^p)$ would have to be modular if it existed. Hence, neither E nor the aforementioned solution (a,b,c) to $x^p + y^p = z^p$ can exist, and Fermat's last theorem holds.

5.6. Looking back and looking forward

The quest to find a proof of Fermat's last theorem lasted more than 350 years, and hundreds of mathematicians tried to attack the problem in many very different ways. It was simply a fantastic challenge that piqued the interest of essentially every mathematician from Fermat to Wiles. Still today, Fermat's last theorem captivates the imagination of math enthusiasts across the world. It is curious, though, that Fermat's last theorem has virtually no interesting consequences other than the statement itself.

However, the study of the solutions of such a simple equation ($x^n + y^n = z^n$) has been the driving force in developing an immense amount of extremely interesting mathematics. The statement of Fermat's last theorem may not have relevant corollaries, but the tools that were used in the proof are incredibly important and offer a vast range of very useful applications.

The final stages of the proof of Fermat's last theorem (as outlined in Section 5.5) represent one of the biggest triumphs of modern mathematics — not just because a 358-year-old problem was solved, but for the fundamental advances in the theory of elliptic curves and modular forms that were produced in order to verify Fermat's claim. This was no small enterprise; we have already briefly described the remarkable involvement of many important mathematicians (Shimura, Taniyama, Weil, Frey, Serre, Ribet, Wiles, Taylor, Breuil, Conrad, and Diamond, among many others). Just the proof of the modularity theorem (Theorem 5.4.6) occupies more than 200 pages of research articles (that's only counting [**Wil95**], [**TW95**] and [**BCDT01**]), and

many books have been written to explain the brilliant mathematics developed for the proof (see **[CSS00]** for a graduate-level textbook).

Fermat's last theorem has been proved, but the broad areas of research that this book touches on (namely algebraic number theory, algebraic geometry and their intersection, arithmetic geometry) have seen an exponential growth over the last couple of centuries, and they continue to grow at a vigorous pace. Nowadays, there is an immense amount of research being done on elliptic curves, modular forms, and generalizations of the modularity theorem to other settings (abelian varieties, elliptic curves over number fields, etc.). Many questions remain unanswered; for instance,

- Are there elliptic curves over $\mathbb{Q}$ of arbitrarily high rank? See Conjecture 2.4.7 and the discussion in the same section.
- Is the Shafarevich-Tate group of an elliptic curve, $Ш(E/\mathbb{Q})$, always a finite group?
- Is the Birch and Swinnerton-Dyer conjecture true for all elliptic curves? See Conjecture 5.2.1 and Section 5.2. The Clay Mathematics Institute has offered a reward of one million dollars for a proof (or counterexample!) of this celebrated conjecture.

These are just three questions of great (huge!) interest to number theorists, but there are many other interesting questions and challenging problems being formulated as the reader stares at this page. The Preface to this book contains a list of suggested reading material so that the reader can continue to learn (more rigorously, and in depth) about elliptic curves, modular forms, and their L-functions.

5.7. Exercises

Exercise 5.7.1. Let $E/\mathbb{Q}$ be an elliptic curve and let $p \geq 2$ be a prime. Define $E^{\mathrm{ns}}(\mathbb{F}_p)$ to be the set of all non-singular points on $E(\mathbb{F}_p)$, and write $N_p^{\mathrm{ns}} = |E^{\mathrm{ns}}(\mathbb{F}_p)|$. For instance, if p is a prime of good reduction, then $E^{\mathrm{ns}}(\mathbb{F}_p) = E(\mathbb{F}_p)$ and $N_p^{\mathrm{ns}} = N_p = p + 1 - a_p$.

Suppose that $E/\mathbb{Q}$ has bad reduction at p. Show that:

$$N_p^{\mathrm{ns}} = \begin{cases} p-1 & \text{if E has split multiplicative reduction at } p; \\ p+1 & \text{if E has non-split multiplicative reduction at } p; \\ p & \text{if E has additive reduction at } p. \end{cases}$$

Conclude that $L_p(p^{-1}) = \frac{N_p^{\mathrm{ns}}}{p}$ for every $p \geq 2$ (including good and bad primes), where the function $L_p(T)$ appears in Definition 5.1.1. (Hint: write $E : f(x,y) = 0$ and express $f(x,y) = ((y-y_0) - \alpha(x-x_0)) \cdot ((y-y_0) - \beta(x-x_0)) - (x-x_0)^3$ where (x_0, y_0) is the singular point for $E(\mathbb{F}_p)$. Exercise 2.12.11 shows that there is (at most) one singular point in $E(\mathbb{F}_p)$, at least for $p \geq 3$.)

Exercise 5.7.2. Prove Proposition 5.1.5. (Hint: $\frac{1}{1-x} = 1 + x + x^2 + \cdots = \sum_{n \geq 0} x^n$, and use the Fundamental Theorem of Arithmetic.)

Exercise 5.7.3. Prove the parity conjecture 5.2.5, assuming the Birch and Swinnerton-Dyer conjecture and the functional equation of $L(E,s)$. (Hint: use the Taylor expansion of $L(E,s)$ around $s = 1$.) Conclude that, if the root number $w(E/\mathbb{Q}) = -1$, then $E(\mathbb{Q})$ is infinite.

Exercise 5.7.4. Let $f(z) = \sum_{n \geq 1} a_n q^n$ be a cusp form in $S_k(\Gamma_0(N))$, and define the Mellin transform of $f(z)$ by

$$\widehat{f}(s) = \int_0^\infty f(iy) y^s \frac{dy}{y}.$$

Show that $\widehat{f}(s) = (2\pi)^{-s}\Gamma(s)L(f,s)$, where $\Gamma(s)$ is the Gamma function and $L(f,s)$ is the L-function attached to f. (You may ignore convergence issues and assume that integrals and infinite sums commute.)

Exercise 5.7.5. Let $p > 3$ be a prime and suppose that a, b, c are pairwise relatively prime integers such that $a^p + b^p = c^p$ and $abc \neq 0$. Let $E/\mathbb{Q}$ be the elliptic curve (Frey curve) defined by

$$E : y^2 = x(x - a^p)(x + b^p).$$

The goal of this exercise is to show that E is semistable with conductor $N_E = \prod_{\ell | abc} \ell$.

(1) Show that, after rearranging a, b and c if necessary, we can assume that $a \equiv 0 \bmod 2$ and $b \equiv c \equiv 1 \bmod 4$. (Hint: if $2|a$ and $b \equiv 3 \bmod 4$, consider $a^p + (-c)^p = (-b)^p$.)

(2) Calculate the discriminant Δ of $E/\mathbb{Q}$.

(3) Show that $E/\mathbb{Q}$ has good reduction at all primes ℓ that do not divide abc.

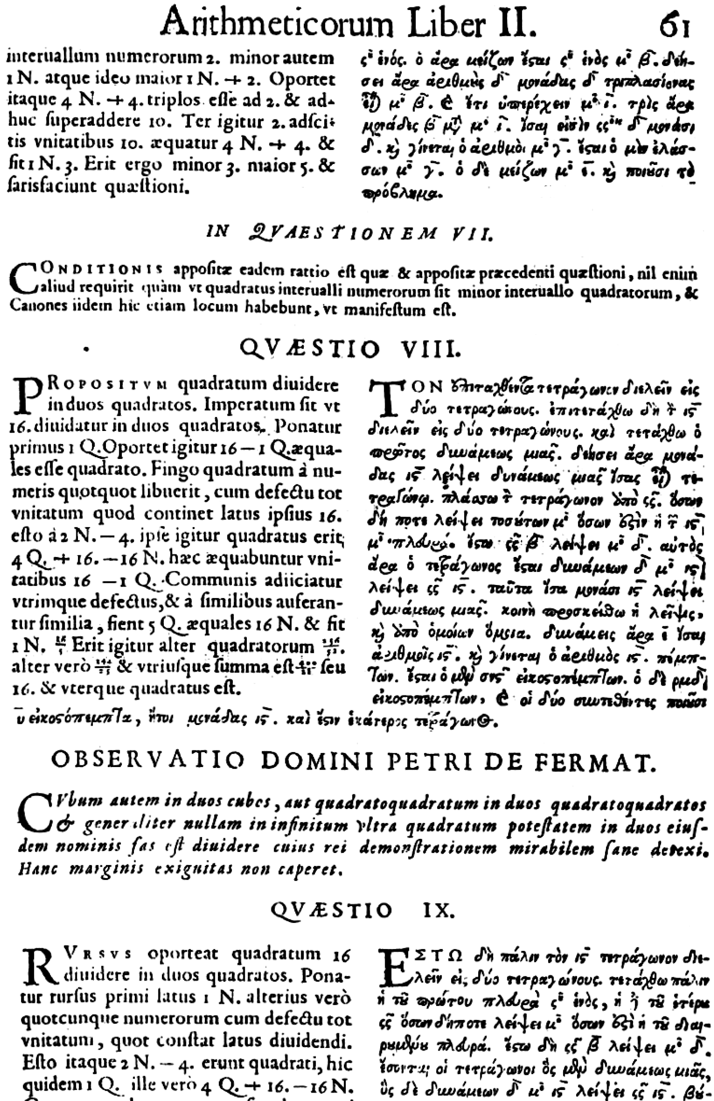
(4) Show that if $\ell \geq 3$ is a prime dividing abc, then $E/\mathbb{Q}$ has bad multiplicative reduction at ℓ.

(5) Show that $E/\mathbb{Q}$ has bad multiplicative reduction at $\ell = 2$. (Hint: use the following change of variables

$$x = \frac{X}{4}, \quad y = \frac{Y}{8} + \frac{3X}{8}$$

to find another model isomorphic to $E/\mathbb{Q}$. Show that this model has coefficients in $\mathbb{Z}$, and analyze the reduction at $\ell = 2$.)

(6) Conclude that the conductor of E is precisely $N_E = \prod_{\ell|abc} \ell$. (See Definition 5.1.7.)



Arithmeticorum Liber II. 61

interuallum numerorum 2. minor autem 1 N. atque ideo maior 1 N. + 2. Oportet itaque 4 N. + 4. triplos esse ad 2. & adhuc superaddere 10. Ter igitur 2. adscitis vnitatibus 10. æquatur 4 N. + 4. & fit 1 N. 3. Erit ergo minor 3. maior 5. & satisfaciunt quæstioni.

ϛ ἑνός. ὁ ἄρα μείζων ἔσται ϛ ἑνὸς μ β. δεήσει ἄρα ἀριθμοὺς δ μονάδας δ τριπλασίονας εἶναι μ β. καὶ ἔτι ὑπερέχειν μ ι. τρὶς ἄρα μονάδες β μετὰ μ ι. ἴσαι εἰσὶν ϛϛ δ μονάσι δ. καὶ γίνεται ὁ ἀριθμὸς μ γ. ἔσται ὁ μὲν ἐλάσσων μ γ. ὁ δὲ μείζων μ ε. καὶ ποιοῦσι τὸ πρόβλημα.

IN QVAESTIONEM VII.

CONDITIONIS appositæ eadem ratio est quæ & appositæ præcedenti quæstioni, nil enim aliud requirit quàm vt quadratus interualli numerorum sit minor interuallo quadratorum, & Canones iidem hic etiam locum habebunt, vt manifestum est.

QVÆSTIO VIII.

PROPOSITVM quadratum diuidere in duos quadratos. Imperatum sit vt 16. diuidatur in duos quadratos. Ponatur primus 1 Q. Oportet igitur 16 − 1 Q. æquales esse quadrato. Fingo quadratum à numeris quotquot libuerit, cum defectu tot vnitatum quod continet latus ipsius 16. esto à 2 N. − 4. ipse igitur quadratus erit 4 Q. + 16. − 16 N. hæc æquabuntur vnitatibus 16 − 1 Q. Communis adiiciatur vtrimque defectus, & à similibus auferantur similia, fient 5 Q. æquales 16 N. & fit 1 N. $\frac{16}{5}$. Erit igitur alter quadratorum $\frac{256}{25}$. alter verò $\frac{144}{25}$. & vtriusque summa est $\frac{400}{25}$. seu 16. & vterque quadratus est.

Τὸν ἐπιταχθέντα τετράγωνον διελεῖν εἰς δύο τετραγώνους. ἐπιτετάχθω δὴ τὸν ιϛ διελεῖν εἰς δύο τετραγώνους. καὶ τετάχθω ὁ πρῶτος δυνάμεως μιᾶς. δεήσει ἄρα μονάδας ιϛ λείψει δυνάμεως μιᾶς ἴσας εἶναι τετραγώνῳ. πλάσσω τὸν τετράγωνον ἀπὸ ϛϛ. ὅσων δή ποτε λείψει τοσούτων μ ὅσων ἐστὶν ἡ τῶν ιϛ μ πλευρά. ἔστω ϛϛ β λείψει μ δ. αὐτὸς ἄρα ὁ τετράγωνος ἔσται δυνάμεων δ μ ιϛ λείψει ϛϛ ιϛ. ταῦτα ἴσα μονάσι ιϛ λείψει δυνάμεως μιᾶς. κοινὴ προσκείσθω ἡ λεῖψις, καὶ ἀπὸ ὁμοίων ὅμοια. δυνάμεις ἄρα ε ἴσαι ἀριθμοῖς ιϛ. καὶ γίνεται ὁ ἀριθμὸς ιϛ πέμπτων. ἔσται ὁ μὲν σνϛ εἰκοστοπέμπτων. ὁ δὲ ρμδ εἰκοστοπέμπτων, καὶ οἱ δύο συντεθέντες ποιοῦσι υ εἰκοστόπεμπτα, ἤτοι μονάδας ιϛ. καὶ ἔστιν ἑκάτερος τετράγωνος.

OBSERVATIO DOMINI PETRI DE FERMAT.

CVbum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos & generaliter nullam in infinitum vltra quadratum potestatem in duos eiusdem nominis fas est diuidere cuius rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.

QVÆSTIO IX.

RVRSVS oporteat quadratum 16 diuidere in duos quadratos. Ponatur rursus primi latus 1 N. alterius verò quotcunque numerorum cum defectu tot vnitatum, quot constat latus diuidendi. Esto itaque 2 N. − 4. erunt quadrati, hic quidem 1 Q. ille verò 4 Q. + 16. − 16 N. Cæterum volo vtrumque simul æquari vnitatibus 16. Igitur 5 Q. + 16. − 16 N. æquatur vnitatibus 16. & fit 1 N. $\frac{16}{5}$ erit

Ἔστω δὴ πάλιν τὸν ιϛ τετράγωνον διελεῖν εἰς δύο τετραγώνους. τετάχθω πάλιν ἡ τοῦ πρώτου πλευρὰ ϛ ἑνός, ἡ δὲ τοῦ ἑτέρου ϛϛ ὅσων δήποτε λείψει μ ὅσων ἐστὶν ἡ τοῦ διαιρουμένου πλευρά. ἔστω δὴ ϛϛ β λείψει μ δ. ἔσονται οἱ τετράγωνοι ὃς μὲν δυνάμεως μιᾶς, ὃς δὲ δυνάμεων δ μ ιϛ λείψει ϛϛ ιϛ. βούλομαι τοὺς δύο λοιπὸν συντεθέντας ἴσους εἶναι μ ιϛ. δυνάμεις ἄρα ε μ ιϛ λείψει ϛϛ ιϛ ἴσαι μ ιϛ. καὶ γίνεται ὁ ἀριθμὸς ιϛ πέμπτων.

H iii

Figure 9. A 1670 edition of Diophantus' *Arithmetica*, which includes the original Greek text, a Latin translation, and Fermat's commentary: "Observatio Domini Petri de Fermat". In this page Fermat states his famous last theorem.

Appendix A

PARI/GP and Sage

This appendix is meant as a brief introduction to the usage of the software packages PARI/GP and Sage, oriented to the study of elliptic curves and modular forms. The websites for these packages are:

- PARI/GP: http://pari.math.u-bordeaux.fr/
- Sage: http://www.sagemath.org/

but *notice* that you can call PARI/GP from Sage, so I would recommend simply installing Sage on your computer. I strongly recommend that you use the "notebook" option in Sage and interact with the software through your favorite internet browser (e.g. Firefox). Sage can also be found online (although the performance, usually, is slower than a local version on your computer):

- Sage online: http://www.sagenb.org/

Both packages have online manuals and specific sections on elliptic curves.

A.1. Elliptic curves

A.1.1. Definition of an Elliptic Curve. An elliptic curve is a plane curve E given by a Weierstrass equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with coefficients $a_1, \ldots, a_6$ in some field F. If the field is of characteristic different from 2 or 3, one can find an easier model of the form

$$y^2 = x^3 + Ax + B.$$

In order to work with elliptic curves using the software packages, we need to define the curves first:

- `GP > E = ellinit`$([a_1, a_2, a_3, a_4, a_6])$
- `Sage > E = EllipticCurve`$([a_1, a_2, a_3, a_4, a_6])$
- or `Sage > E = EllipticCurve`$([A, B])$.

Once we have defined an elliptic curve E, we can calculate basic quantities such as the discriminant, the j-invariant or any of the coefficients b_i or c_j (as defined in **[Sil86]**, Ch. III, §1):

- In `GP`, type `E.disc`, `E.c4` or `E.j`,
- In `Sage`, type `E.discriminant()`, `E.c4()` or `E.j_invariant()`.

If the elliptic curve is given by a model of the form $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ but you would rather have a model $y^2 = x^3 + Ax + B$, use the command `E.integral_short_weierstrass_model()`.

Remark A.1.1. Perhaps **the two most useful Sage tricks** are the "Tab" key after an object and "?" after a command to get help. For instance, if we have defined an elliptic curve `E`, then typing

```
E.
```

followed by the "Tab" key displays all possible commands that one can use with an elliptic curve. This is very useful when we do not remember the exact syntax or we are wondering if Sage is capable of doing some particular operation on `E`. Similarly, if we want to know more about the usage of a particular command, then "`E.command_name?`" will display a help box. For example, if we input `E.discriminant?` then Sage tells us that this command returns the discriminant of `E` and provides a couple of examples for the user.

A.1.2. Basic operations. Let us start by using the addition on an elliptic curve. Let E be the curve given by $Y^2 = X^3 + 1$, and suppose we have initialized E as above. This curve has points $P = [0, 1]$ and

$Q = [-1, 0]$. Let us find $P + Q$ and $2P$ (the answers are $[2, -3]$ and $[0, -1]$ respectively). The commands are:

- In GP, the commands are `elladd(E,[0,1],[-1,0])` and, in order to find $2P$, one types `ellpow(E,[0,1],2)`;
- Sage: First we create points on the curve: `P = E([0,1]); Q = E([-1,0])` and now we can do addition: type `P+Q` and `P+P`, or calculate multiples by typing `2*P`, `3*P`, etc.

Notice that Sage will transform affine points to projective coordinates (e.g., `P = E([0,1])` returns `(0 : 1 : 1)` in Sage). If you want to find points on a curve (up to a given bound B on the height of the point), use `E.point_search(B)` in Sage.

A.1.3. Plotting. Here is an example of a 2D-plot with Sage:

```
E = EllipticCurve([0,0,0,0,1]);
Ep = plot(E, -1,2.5,thickness=2);
p1=(2,3); p2=(0,1); p3=(-1,0); p4=(0,-1); p5=(2,-3);
L1=line([p1,p3],rgbcolor=(1,0,0));
L2=line([p5,p3],rgbcolor=(1,0,0));
L3=line([p4,p3],rgbcolor=(1,0,0));
L4=line([p2,p5],rgbcolor=(1,0,0));
L5=line([p4,p1],rgbcolor=(1,0,0));
T1=text('P',[2,3.5]); T2=text('2P',[0.15,1.5]);
T3=text('3P',[-1,.5]); T4=text('4P',[0.15,-1.5]);
T5=text('5P',[2,-3.5]);
P=point([p1,p2,p3,p4,p5],pointsize=30,
rgbcolor=(0,0,0));
PLOT=Ep+T1+T2+T3+T4+T5+L1+L2+L3+L4+L5+P; show(PLOT)
```

The result is the graph that appears in Figure 2. The following is an alternative way to plot points on a curve:

```
Q = E(2,3);
Qplot = plot(Q, pointsize=30)+plot(2*Q, pointsize=30);
show(Qplot)
```

A.1.4. Good and bad reduction. Given a prime p and an elliptic curve $E/\mathbb{Q}$ given by a Weierstrass equation with integer coefficients, we can consider E as a curve over $\mathbb{Z}/p\mathbb{Z}$. The primes that divide the (minimal) discriminant are called bad primes or primes of bad reduction. In Sage, you can find the minimal model of an elliptic curve E by typing `E.minimal_model()`. For example, in Sage, the commands

```
E=EllipticCurve([0,5,0,0,35]);
prime_divisors(E.discriminant())
```

will return `[2,5,7,17]`. You may also use

```
factor(E.discriminant()).
```

Then one can use the command `kodaira_type()` to find out the precise type of reduction: `I0` is good reduction; `Ij`, where $j > 0$ is some positive number, means bad multiplicative reduction; `II, III, IV` or `Ij*`, for $j \geq 0$, or `II*,III*,IV*` mean additive reduction. For an explanation of the terminology of Kodaira symbols, see [**Sil86**], Appendix C, §15. For our example $E : y^2 = x^3 + 5x^2 + 35$, we obtain

`E.kodaira_type(2)` returns `II` (i.e., additive);
`E.kodaira_type(5)` returns `II` (i.e., additive);
`E.kodaira_type(7)` returns `I1` (i.e., multiplicative);
`E.kodaira_type(17)` returns `I2` (i.e., multiplicative);
`E.kodaira_type(11)` returns `I0` (i.e., good).

Note: if the equation is not minimal, some of the prime divisors of the discriminant may not be bad after all. For example,

> `E=EllipticCurve([0,0,0,0,15625]);`
>
> `prime_divisors(E.discriminant())` returns `[2,3,5]` but
>
> `E.kodaira_type(5)` returns `I0` (i.e., good).

This happened because the model $y^2 = x^3 + 15625$ is not minimal ($15625 = 5^6$); we should have used $y^2 = x^3 + 1$ instead.

If $E/\mathbb{Q}$ has good reduction at p, then E defines an elliptic curve over the finite field $\mathbb{Z}/p\mathbb{Z}$ and we can count the number of points modulo p (always including the extra point at infinity). N_p denotes this number of points while $a_p = p + 1 - N_p$. In GP, the command `ellap(E,p)` returns the coefficient a_p and `ellan(E,n)` returns an array with the first n coefficients a_k for $k = 1, \ldots, n$.

In Sage, the command `E.ap(p)` returns a_p while `E.an(n)` returns the nth coefficient (and only the nth), and `E.anlist(n)` provides a list of all the coefficients up to a_n. In Sage you can also directly find the number N_p by typing `E.Np(p)`.

The conductor of $E/\mathbb{Q}$ is another associated quantity that is very useful in practice:

- In Sage, type `E.conductor()`,
- In GP, type `ellglobalred(E)`.

The command `ellglobalred(E)` returns an array [conductor, global minimal model, product of local Tamagawa numbers]. In Sage, you can find a minimal model of an elliptic curve E by typing the command `E.minimal_model()`.

A.1.5. The torsion subgroup. It follows from the Mordell-Weil theorem that the torsion subgroup of an elliptic curve (over a number field) is a finite abelian group. Over $\mathbb{Q}$, a theorem of B. Mazur says that the torsion subgroup is one of the following: $\mathbb{Z}/n\mathbb{Z}$ with $1 \leq n \leq 10$ or $n = 12$, or $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2m\mathbb{Z}$ with $1 \leq m \leq 4$. One can compute the torsion subgroup as follows. The computation is easy, due to a theorem of Nagell and Lutz:

- In GP, the output of `elltors(E)` is a vector `[t, [n, m], [P,Q]]`, where t is the size of the torsion subgroup, which is isomorphic to $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$, generated by the points P and Q. If P is a torsion point, the command `ellorder(P)` provides the order of the element.
- In Sage, `E.torsion_order()` returns the order of the group, while `G = E.torsion_subgroup()` returns the group itself. Then `G.0` and `G.1` return generators for G.

Remark A.1.2. Even though the Nagell-Lutz theorem provides a simple algorithm to calculate the torsion subgroup of an elliptic curve, this method may not be very effective (at least when the discriminant is divisible by many primes). In general, there are better algorithms (for example, see [**Dou98**]).

A.1.6. The free part and the rank. It also follows from the Mordell-Weil theorem that the free part (here *free* is the opposite of torsion) of the group of points $E(K)$ on an elliptic curve (again over a number field K) is generated by a finite number of points P_1, P_2, ..., P_R of infinite order. The number R of generators (of infinite order) is called the rank of $E(K)$. There is no known algorithm that will always terminate and provide the rank and a set of generators. However, the so-called "descent algorithm" will terminate in certain cases (the descent procedure is an algorithm if Ш is finite, and we conjecture that Ш is always finite). The following commands compute lower and upper bounds for the rank and, in some cases, if they coincide, provide the rank of the curve. There are also commands to calculate generators; however, in many situations, the resulting points will only generate a group of finite index in $E(K)$ (the software will warn you when this may be the case). Some of the algorithms take an optional argument of a bound B.

In Sage, the command `E.selmer_rank_bound()` gives an upper bound of the rank, and `E.rank()`, `E.gens()` *try* to find, respectively, the rank and generators modulo torsion... but the computer may not succeed! When these commands are called, Sage is using an algorithm of Cremona in the background (see [**Cre97**]).

A.1.7. Heights and independence. In order to determine if a set of rational points is algebraically independent, we use a pairing arising from the canonical height. The following commands calculate the global Néron-Tate canonical height of a rational point P on a curve E:

In GP use `ellheight(E,P)`;

In Sage simply use `P.height()`, where `P` is a point on E.

If $S = \{P_1, \ldots, P_n\}$ is a set of rational points, we can test whether they are independent using the canonical height matrix. The height pairing of P and Q is defined by $\langle P, Q\rangle = h(P+Q) - h(P) - h(Q)$, where h is the canonical height on E. The height matrix relative to S is a matrix H whose coordinate ij is given by $\langle P_i, P_j\rangle$. The canonical height is a positive definite quadratic form on $E(\mathbb{Q})$ tensored with the reals. Thus, the determinant of H is non-zero if and only if the points in S are independent modulo torsion.

In GP use `S = [P1,P2,P3];`

```
H=ellheightmatrix(E,S); matdet(H);
```

In Sage use `E.height_pairing_matrix([P1,P2,P3])`,

where `P1, P2, P3` are points on E (previously defined). In GP, if `matdet(H)` returns 0, one can calculate generators for the kernel of H with `matker(H)`. Each element of the kernel represents a linear combination of points that adds up to a torsion point. In Sage, you may use `H.kernel()` for the same purpose.

A.1.8. Elliptic curves over $\mathbb{C}$. The period lattice of an elliptic curve $E/\mathbb{Q}$ can be found by typing

```
L=E.period_lattice()
```

and a basis for the period lattice is found simply using `L.basis()`. Using PARI/GP, one can start from a lattice and obtain the associated elliptic curve, as follows:

`L=[1,I];`
`elleisnum(L,4)` returns $G_4(L)$,
  which equals `2268.8726415...`,
`elleisnum(L,6)` returns $G_6(L)$,
  which equals `-3.97...E-33`, i.e., 0,
thus, L corresponds to an elliptic curve
$y^2 = x^3 - (34033.089\ldots)x$.

The elliptic curve $y^2 = x^3 - (34033.089\ldots)x$ is isomorphic to $E/\mathbb{Q}$: $y^2 = x^3 - x$ over $\mathbb{C}$. Thus, $\mathbb{C}/\langle 1, i\rangle \cong E(\mathbb{C})$.

A.2. Modular forms

In this section, all commands we list are to be used in the Sage environment.

A.2.1. The modular group and congruence subgroups. The modular group and main congruence subgroups, defined for any $N > 0$ by

$$\begin{aligned}
\mathrm{SL}(2,\mathbb{Z}) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,c,d \in \mathbb{Z},\ ad-bc=1 \right\}, \\
\Gamma_0(N) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) : c \equiv 0 \bmod N \right\}, \\
\Gamma_1(N) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) : a \equiv d \equiv 1 \bmod N \right\},
\end{aligned}$$

may be defined in Sage using `SL2Z`, `Gamma0(N)`, and `Gamma1(N)`, respectively. Alternatively, $\mathrm{SL}(2,\mathbb{Z})$ can also be defined as $\Gamma_0(1)$. Notice that those 2×2 matrices that define elements of congruence subgroups are stored in Sage as 4-dimensional row vectors. One can use the subcommand `.gens()` on any of the modular and congruence groups to find a set of matrices that generate (multiplicatively) the given group.

You can call the generators by using the suffix `[0]`, `[1]`, etc. Here are some examples:

`A = SL2Z([1,1,0,1]);`

`G = SL2Z.gens()` returns two matrices

$$G[0] = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad G[1] = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

`H = Gamma0(3).gens()` returns six matrices

$$H[0] = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, H[1] = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, H[2] = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

$$H[3] = \begin{pmatrix} 1 & -1 \\ 3 & -2 \end{pmatrix}, H[4] = \begin{pmatrix} 2 & -1 \\ 3 & -1 \end{pmatrix}, H[5] = \begin{pmatrix} -2 & 1 \\ -3 & 1 \end{pmatrix}.$$

The genus of the modular curve $X_0(N)$ can be computed with the command `Gamma0(N).genus()`. Similarly, `Gamma(N).genus()` and `Gamma1(N).genus()` return the genus of $X(N)$ and $X_1(N)$, respectively.

A.2.2. Vector spaces of modular forms. Let Γ be a congruence subgroup of $\mathrm{SL}(2,\mathbb{Z})$ and define:

- $M_k(\Gamma)$, the $\mathbb{C}$-vector space of all modular forms for Γ of weight k;
- $S_k(\Gamma)$, the $\mathbb{C}$-vector space of all cusp forms for Γ of weight k.

Suppose you have already defined a congruence subgroup `G` (for example, `G = Gamma0(3)`) and are interested in forms of weight `k`. The vector spaces of modular forms and cusp forms can be defined in Sage by

`M=ModularForms(G,k)` or `ModularForms(G,k,prec=m)`
if you want q-series expansions up to q^m;

`S=CuspForms(G,k)` or `CuspForms(G,k,prec=m)`.

The precision is set to 6 by default. If you want to find the dimension or a basis, you can use the suffix `.dimension()` or `.basis()`,

respectively. Here is an example:

`M=ModularForms(Gamma0(3),4, prec=10);`

`M.dimension()` returns `2`;

`M.basis()` returns the forms:

$[1 + 240q^3 + 2160q^6 + 6720q^9 + O(q^{10}),$

$q + 9q^2 + 27q^3 + 73q^4 + 126q^5 + 243q^6$

$+344q^7 + 585q^8 + 729q^9 + O(q^{10})].$

The command `CuspForms(Gamma0(3),4,prec=10)` returns only the 0 vector space. Notice that even though the modular form $q + 9q^2 + 27q^3 + O(q^4)$ vanishes at the cusp at infinity (because $a_0 = 0$ in the expansion), it is not a cusp form for $\Gamma_0(3)$ because it does not vanish at *all* the cusps of $X_0(3)$ (infinity is not the only cusp!). The command `AllCusps(N)` produces a list of all (representatives of) cusps of $X_0(N)$.

`AllCusps(3)` returns `[(inf),(0)]`.

A.3. L-functions

Let $E/\mathbb{Q}$ be an elliptic curve, and let $L(E, s)$ be the Hasse-Weil L-function associated to E, as in Definition 5.1.1. This L-function is defined in Sage using the command

```
L=E.lseries()
```

or one can use `L=E.lseries().dokchitser()` to use Dokchitser's algorithms to calculate values ([**Dok04**]). Once we have defined $L = L(E, s)$, we can evaluate L. For example:

`E=EllipticCurve([1,2,3,4,5]);`

`L=E.lseries();`

`L(1)` which returns 0,

`L(1+I) = -0.485502124065793 + 0.627256178203893*I.`

The value $L(E, 1) = 0$ is predicted in this case by the Birch and Swinnerton-Dyer conjecture (Conjecture 5.2.1), since the rank of E is > 0 (in fact, the rank is 1). One can also plot $L(E, x)$ when x takes

real values (because $L(E,x)$ is real valued for $x \in \mathbb{R}$). For instance, the graph in Figure 2 was created with the following lines of code:

```
E0=EllipticCurve([0,-1,1,-10,-20]);
L0=E0.lseries().dokchitser();
P0=plot(lambda x:  L0(x).real(),0, 3);
show(P0,xmin=-0.5, ymin=0, dpi=150).
```

If you want to create a PDF file with your graph, you can use

```
P=plot(lambda x:  real(L0(x)),0, 3).save(
"bsdrank0.pdf",xmin=-0.5, ymin=-0.2, dpi=150).
```

You may also want to calculate the Taylor polynomial of $L(E,s)$ around the point $x=a$ of degree $n-1$ with `L.taylor_series(a,n)`.

A.3.1. Data related to the BSD conjecture. The Shafarevich-Tate group of $E/\mathbb{Q}$ is defined in Sage by `E.sha()` but, in general, it is difficult to calculate its order. The user can calculate a conjectural value of Sha by typing `E.sha().an()`. The conductor N of $E/\mathbb{Q}$ is calculated with `E.conductor()`. The Tamagawa product $\prod_{p|N} c_p$ can be calculated directly with `E.tamagawa_product()` or the individual Tamagawa numbers c_p, for each prime $p|N$, may be calculated with `E.tamagawa_number(p)`. The regulator of $E/\mathbb{Q}$ can be calculated by `E.regulator()`. Finally, the real period Ω_E is calculated as follows:

```
E=EllipticCurve([1,2,3,4,5]);
M=E.period_lattice();
```

Then `M.omega` returns $\Omega_E = 2.78074001376673\ldots$.

The reader should try to use the commands above to calculate all the invariants listed in Examples 5.2.3 and 5.2.4 (see Figure 6 and Figure 7).

A.4. Other Sage commands

- Continued fractions:

`continued_frac_list(N)` returns the continued fraction of N; `continued_frac_list(N,partial_convergents=True)` or `convergents(v)` return convergents for the cont. frac. v.

- The Kronecker symbol (defined in Example 1.3.3):

`kronecker(-n,m)` returns the Kronecker symbol $\left(\frac{-n}{m}\right)$.

Appendix B

Complex analysis

In this appendix we review some of the basic notions of complex numbers and the theory of analytic and meromorphic functions on the complex plane. This brief appendix is by no means a replacement for a good course or a good book on complex analysis such as [**Ahl79**].

B.1. Complex numbers

The complex numbers, usually denoted by $\mathbb{C}$, are defined as an extension of the real numbers $\mathbb{R}$. Over the reals, the equation $x^2 + 1 = 0$ has no solutions, so we define a new number i that satisfies $i^2 = -1$. Therefore $x^2 + 1 = 0$ now has two solutions, namely i and $-i$. We define $\mathbb{C}$ by adjoining our new number i to $\mathbb{R}$:

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R},\ i^2 = -1\}.$$

The real and imaginary parts of a complex number $\alpha = a + bi$ are denoted, respectively, by $\Re(\alpha) = a$ and $\Im(\alpha) = b$. If $\Im(\alpha) = b = 0$ we say that α is a *real number*, and if $\Re(\alpha) = a = 0$ we say that α is *purely imaginary*. We can add and multiply two complex numbers $\alpha = a+bi$ and $\beta = c+di$ to obtain a new complex number, as follows:

$$\begin{aligned} \alpha + \beta &= (a + bi) + (c + di) = (a + c) + (b + d)i; \text{ and} \\ \alpha \cdot \beta &= (a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i. \end{aligned}$$

The set of all complex numbers together with the operations of addition and multiplication form a field (see Exercise B.7.1).

There are two other operations on complex numbers that occur often: complex conjugation and calculating the modulus, or absolute value. The *complex conjugate* of $\alpha = a+bi$ is $\overline{\alpha} = a-bi$. The *modulus* or *absolute value* of α is

$$|\alpha| = \sqrt{\alpha \cdot \overline{\alpha}} = \sqrt{(a+bi)(a-bi)} = \sqrt{a^2+b^2}.$$

Notice that, for any $\alpha, \beta \in \mathbb{C}$, we have

$$\overline{\alpha+\beta} = \overline{\alpha}+\overline{\beta},\ \overline{\alpha \cdot \beta} = \overline{\alpha} \cdot \overline{\beta},\ |\alpha\beta| = |\alpha||\beta|, \text{ and } |\alpha+\beta| \leq |\alpha|+|\beta|.$$

We constructed the complex numbers by adjoining i to $\mathbb{R}$ so that $i \in \mathbb{C}$ and therefore the equation $x^2+1=0$ has two solutions in $\mathbb{C}$. But something extremely surprising happened in this construction. It turns out that not only x^2+1 has a root in $\mathbb{C}$ but, in fact, *every polynomial* with complex coefficients has a root in $\mathbb{C}$. This is an extremely important result:

Theorem B.1.1 (Fundamental Theorem of Algebra)**.** *Let $p(z)$ be a polynomial*

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0$$

with complex coefficients $a_i \in \mathbb{C}$ and degree ≥ 1. Then there exists a complex number $\alpha \in \mathbb{C}$ such that $p(\alpha) = 0$.

The proof is left to the reader (Exercise B.7.6).

B.2. Analytic functions

Definition B.2.1. Let $\alpha \in \mathbb{C}$ and $\delta \in \mathbb{R}^+$. An *open disc* $D_\delta(\alpha)$ in the complex plane, centered at α and of radius $\delta > 0$, is the set

$$D_\delta(\alpha) = \{z \in \mathbb{C} : |z-\alpha| < \delta\}.$$

Definition B.2.2. We say that a set $S \subseteq \mathbb{C}$ is *open* if for every $\alpha \in S$ there is a real number $\delta > 0$ such that $D_\delta(\alpha) \subseteq S$. We say that a set $T \subseteq \mathbb{C}$ is *closed* if the complement of T in $\mathbb{C}$, i.e., $\mathbb{C} - T$, is open.

Definition B.2.3. A non-empty connected open set in the complex plane is called a *region*.

Let U be a region in the complex plane and let $f(z) : U \to \mathbb{C}$ be a complex-valued function on U. Let $\alpha \in U$. We say that f *has a derivative at* α if the usual limit converges:

$$f'(\alpha) = \lim_{h \to 0} \frac{f(\alpha + h) - f(\alpha)}{h},$$

where h runs over complex numbers inside U that approach 0. Alternatively (or more precisely), we can define $f'(z)$ using ϵ and δ as follows. We say that f has a derivative at α with value $m = f'(\alpha)$ if the following statement holds: for every real $\epsilon > 0$ there exists a real $\delta > 0$ such that, if $h \in D_\delta(\alpha)$, then

$$\left| \frac{f(\alpha + h) - f(\alpha)}{h} - m \right| < \epsilon.$$

Definition B.2.4. Let $U \subseteq \mathbb{C}$ be a region and let $f(z)$ be a complex-valued function $f : U \to \mathbb{C}$ defined for every $z \in U$. We say that $f(z)$ is *analytic* (or *holomorphic*, or *entire*) on U if it has a derivative at each $z \in U$.

Example B.2.5. The function $f(z) = z$ is analytic on the whole complex plane $\mathbb{C}$ (Exercise B.7.3). The function $g(z) = 1/z$ is analytic on $\mathbb{C} - \{0\}$.

It is not hard to show that the sum, product and composition of two analytic functions are also analytic. Thus, all polynomials in one variable with complex coefficients define analytic functions. Similarly, the quotient of two analytic functions is analytic except at the zeros of the denominator. Thus, all rational functions (quotients of polynomials) are analytic in the complex plane except at the zeros of the polynomial in the denominator.

Remark B.2.6. Let U be a region of $\mathbb{C}$ and let $f : U \to \mathbb{C}$ be an analytic function. We write $f(z)$ where $z = x + yi \in U$, with $x, y \in \mathbb{R}$. We may also write

$$f(z) = u(z) + v(z)i,$$

where $u, v : U \to \mathbb{R}$ are real-valued functions. Since f is analytic on U, the functions f, u and v are continuous on U (Exercise B.7.4). Since f is analytic, the limit

$$f'(z) = \lim_{h \to 0} \frac{f(z + h) - f(z)}{h} \tag{B.1}$$

exists for every $z \in U$. The parameter h runs over complex numbers in U approaching zero, but we may restrict h to real values (thus, we are calculating $\partial f/\partial x$). The value of the limit in Eq. (B.1) does not change under this restriction, and this means that the partial derivative of f with respect to x equals $f'(z)$. Hence

$$f'(z) = \frac{\partial f}{\partial x} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x} i.$$

Similarly, we may restrict h to purely imaginary values $h = ik$, and then

$$\begin{aligned} f'(z) &= \lim_{h\to 0} \frac{f(z+h)-f(z)}{h} = \lim_{k\to 0} \frac{f(z+ik)-f(z)}{ik} \\ &= \frac{1}{i} \cdot \lim_{k\to 0} \frac{f(z+ik)-f(z)}{k} = (-i)\frac{\partial f}{\partial y}. \end{aligned}$$

It follows that $f'(z) = (-i)\frac{\partial f}{\partial y} = (-i)\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}$. Therefore,

$$f'(z) = \frac{\partial f}{\partial x} = (-i)\frac{\partial f}{\partial y} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x} i = \frac{\partial v}{\partial y} - \frac{\partial u}{\partial y} i.$$

The last equality implies that the real and imaginary parts of every analytic function must satisfy the following differential equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{B.2}$$

These are called the *Cauchy-Riemann differential equations.*

Differentiability (or being analytic) over $\mathbb{C}$, as in Definition B.2.4, is a much stronger condition than differentiability over $\mathbb{R}$. Indeed, the existence of a complex derivative implies that the function is in fact *infinitely differentiable* and *locally equal to its own Taylor series.* We explain what these terms mean in the following theorem.

Theorem B.2.7. *Let $U \subseteq \mathbb{C}$ be a region and let $f : U \to \mathbb{C}$ be analytic. Then f has derivatives of all orders on U (i.e., the derivatives $f'(z), f''(z), \ldots$ and, more generally, $f^{(n)}(z)$ for all $n \geq 1$ are continuous and differentiable complex-valued functions on U).*

Moreover, for every $\alpha \in U$, the Taylor series of $f(z)$ about $z = \alpha$ converges to $f(z)$ in some neighborhood of α. In other words, for

every $\alpha \in U$, there is a real $\delta > 0$ such that the Taylor series

$$T(z;\alpha) = \sum_{n=0}^{\infty} \frac{f^{(n)}(\alpha)}{n!}(z-\alpha)^n$$

converges for all $z \in D_\delta(\alpha)$, and $T(z;\alpha) = f(z)$.

Conversely, if $S(z) = \sum_{n=0}^{\infty} a_n(z-\alpha)^n$ is a power series with complex coefficients a_n with a radius of convergence R (i.e., $S(z)$ converges for all $z \in \mathbb{C}$ with $|z-\alpha| < R$), then $S(z)$ defines an analytic function on the open disc $D_R(\alpha)$.

Example B.2.8. Let $f(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$. The radius of convergence of this series is infinite (over $\mathbb{C}$ as well as over $\mathbb{R}$), so it defines an analytic function in the complex plane. The function $f(z)$ is, of course, the complex exponential function which we discuss below in B.4 in some more detail. Similarly, we define $\sin(z)$ and $\cos(z)$ using the usual Taylor expansions

$$\sin(z) = \sum_{n=0}^{\infty}(-1)^{2n+1}\frac{z^{2n+1}}{(2n+1)!}, \quad \cos(z) = \sum_{n=0}^{\infty}(-1)^{2n}\frac{z^{2n}}{(2n)!}.$$

Since the radius of convergence of these series is infinite, $\sin(z)$ and $\cos(z)$ define analytic functions on $\mathbb{C}$.

B.3. Meromorphic functions

At this juncture, it is useful to extend the complex numbers by introducing a point at infinity ∞. We will write $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ for the *extended complex plane*. We set the convention that every straight line shall pass through the point at infinity. (Note that $\widehat{\mathbb{C}}$ is simply the projective line over $\mathbb{C}$, i.e., $\mathbb{P}^1(\mathbb{C})$. See Appendix C for an introduction to the projective line and projective geometry.)

With this definition of ∞, suppose that $f(z)$ is a complex-valued function not defined at α. The expression

$$\lim_{z\to\alpha} f(z) = \infty$$

means that $|f(z)|$ is unbounded as z approaches α. For instance, $f(z) = 1/z$ is not defined at 0 and $\lim_{z\to 0} 1/z = \infty$. This "$\infty$" is the complex point at infinity, and it should not be confused with the

infinity that we use in real analysis ("very, very far along the positive x-axis"). In fact, in $\mathbb{R}$, the limit $\lim_{x\to 0} 1/x$ is undefined (as the value may be $\pm\infty$ depending on how we approach 0), but in $\widehat{\mathbb{C}}$, the limit $\lim_{z\to 0} 1/z = \infty$ simply means that if z is close to 0, then $1/z$ is far from 0 (in some direction, not necessarily along the x-axis).

Suppose that $f(z)$ is some complex-valued function that is not defined at α but is analytic in a neighborhood of α. How can f fail to be analytic at α? The function $f(z)$ may have a *removable singularity* (e.g., $\sin(z)/z$), an *essential singularity* (e.g., $\sin(1/z)$) or a *pole* (e.g., $1/z$). Here we will only discuss poles in some detail (for a complete discussion, see **[Ahl79]**, Ch. 4, §3).

Definition B.3.1. Let f be a complex-valued function, and let $\alpha \in \mathbb{C}$. We say that f has a *pole* (or *isolated pole*) at $z = \alpha$ if:

(1) The function $f(z)$ is analytic on some disc $D_\delta(\alpha)$ centered at α, except at α itself. In other words, f is analytic on the punctured disc

$$\{z \in \mathbb{C} : 0 < |z - \alpha| < \delta\}$$

for some $\delta > 0$; and

(2) The limit of f at α is infinite:

$$\lim_{z\to\alpha} f(z) = \infty.$$

Definition B.3.2. A function $f(z)$ is *meromorphic* in a region U if f is analytic on U except for a set of isolated poles.

Remark B.3.3. Suppose that $f(z)$ is meromorphic in a region U with an isolated pole at $\alpha \in U$. It does not make sense to write $f : U \to \mathbb{C}$, since $\lim_{z\to\alpha} f(\alpha) = \infty$. Instead, we may write $f : U \to \widehat{\mathbb{C}}$.

Example B.3.4. Let $p(z)$ and $q(z)$ be polynomials in $\mathbb{C}[z]$ such that p and q have no common factors. Then the rational function $p(z)/q(z)$ is a meromorphic function with isolated poles at the zeros of $q(z)$.

Example B.3.5. The function $\sin(1/z)$ has infinitely many zeros accumulating near $z = 0$ (there is a zero at each $z = 1/(\pi k)$ for each $k \geq 1$). Therefore, $g(z) = (\sin(1/z))^{-1}$ is not meromorphic because the singularity at 0 is not isolated. In fact, the function g

has infinitely many poles in any open neighborhood of 0. Notice, however, that $(\sin(z))^{-1}$ is a meromorphic function.

Remark B.3.6. Let $f(z)$ be a function that is analytic in a disc $D_R(\alpha)$ except, perhaps, at $\alpha \in \mathbb{C}$. Then $f(z)$ has a Laurent expansion of the form

$$f(z) = \sum_{n=-\infty}^{\infty} c_n(z-\alpha)^n.$$

Then, the function $f(z)$:

(1) is analytic at α if $c_n = 0$ for all $n < 0$ and $f(\alpha) = c_0$ (if $f(\alpha) \neq c_0$, or if $f(\alpha)$ is undefined, then there is a removable singularity at α),

(2) is meromorphic at α if there is some $M > 0$ such that $c_n = 0$ for all $n < -M$; i.e., the expansion of $f(z)$ is of the form

$$f(z) = \sum_{n=-M}^{\infty} c_n(z-\alpha)^n,$$

and

(3) has an essential singularity at α if there are infinitely many $n < 0$ such that $c_n \neq 0$.

B.4. The complex exponential function

The usual real exponential function e^x can be extended to the field of complex numbers as follows. Let $z = x + yi$ with $x, y \in \mathbb{R}$. Then we define e^z by

$$e^z = e^{x+yi} := e^x(\cos(y) + \sin(y)i).$$

Equivalently, e^z can be defined as a Taylor series (which coincides with the Taylor series of the real valued exponential function):

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

If $z = x + yi$ with $x, y \in \mathbb{R}$:

$$\begin{aligned} |e^z| &= |e^{x+yi}| = |e^x \cos(y) + (e^x \sin(y))i| \\ &= \sqrt{(e^x \cos(y))^2 + (e^x \sin(y))^2} \\ &= \sqrt{e^{2x}(\cos^2(y) + \sin^2(y))} = e^x. \end{aligned}$$

Notice that, if $\theta \in \mathbb{R}$, then $e^{\theta i}$ is a complex number that lies on the unit complex circle $\{z \in \mathbb{C} : |z| = 1\}$. Indeed, by the formula above, $|e^{\theta i}| = |e^{0+\theta i}| = e^0 = 1$.

In the theory of L-functions, we often calculate powers of natural numbers $n \in \mathbb{N}$ with complex exponents $s \in \mathbb{C}$. Next, we define what n^s means precisely. If $n \in \mathbb{N}$ and $s = x + yi \in \mathbb{C}$, we define $n^s = e^{\log(n)s}$, i.e.,

$$\begin{aligned} n^s &= e^{\log(n)s} = e^{\log(n)x + \log(n)yi} \\ &= e^{\log(n)x}(\cos(\log(n)y) + \sin(\log(n)y)i) \\ &= n^x(\cos(\log(n)y) + \sin(\log(n)y)i). \end{aligned}$$

B.5. Theorems in complex analysis

In this section we state some of the most important and useful theorems about analytic functions. We have already stated two fundamental theorems, namely Theorems B.1.1 and B.2.7.

The first two theorems concern line integrals along closed curves. If γ is a closed curve (the starting point is equal to the end point) in $\mathbb{C}$, and $f(z)$ is a function defined at every point of γ, then the symbol $\int_\gamma f(z)dz$ represents the line integral of $f(z)$ along γ. A curve is contractible in a region U if it can be continuously shrunk to a point, always staying inside U. The winding number of a curve γ with respect to a point $\alpha \in \mathbb{C}$, denoted by $n(\gamma, \alpha)$, counts the number of times that the path γ winds around α. The winding number is positive if the curve goes around α in the counterclockwise direction, and negative otherwise. (See **[Ahl79]**, Ch. 4.)

Theorem B.5.1 (Cauchy's Theorem)**.** *Let U be a region in $\mathbb{C}$, let $f(z)$ be a complex-valued function that is analytic on U, and let γ be*

any contractible closed curve contained in U. Then

$$\int_\gamma f(z)dz = 0.$$

Theorem B.5.2 (Cauchy's Integral Formula)**.** *Let $f(z)$ be a function that is analytic in a region U, and let γ be a closed curve inside U. For any point α not on γ, we have*

$$n(\gamma, \alpha) \cdot f(\alpha) = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{z - \alpha} dz,$$

where $n(\gamma, \alpha)$ is the winding number of γ around α.

Cauchy's Theorem B.5.1 has the following converse.

Theorem B.5.3 (Morera's Theorem)**.** *If $f(z)$ is defined and continuous in a region $U \subseteq \mathbb{C}$, and if $\int_\gamma f(z)dz = 0$ for all closed curves γ in U, then $f(z)$ is analytic in U.*

Another important theorem about line integrals is the Residue Theorem (see **[Ahl79]**, Ch. 4, §5.1). Before stating the next theorem, we remind the reader that a region is by definition a non-empty connected open set.

Theorem B.5.4 (The Maximum Principle)**.** *If $f(z)$ is analytic and non-constant in a region U, then its absolute value $|f(z)|$ has no maximum in U. Alternatively, if $f(z)$ is an analytic function on a closed bounded set T, then the maximum of $|f(z)|$ occurs on the boundary of T.*

Theorem B.5.5 (Liouville's Theorem)**.** *A function which is analytic and bounded in the whole complex plane must be constant.*

We say that α in a set $S \subseteq \mathbb{C}$ is an *accumulation point* in S if for every $\delta > 0$ there is point $\beta \in S$, $\beta \neq \alpha$ such that $|\beta - \alpha| < \delta$.

Theorem B.5.6. *If $f(z)$ and $g(z)$ are analytic in a region U, and if $f(z) = g(z)$ for every z in a set S which has an accumulation point in U, then $f(z)$ is identically equal to $g(z)$ on all points of U.*

The previous theorem has some remarkable consequences: if $f(z)$ is analytic in U and it is identically zero in a set $S \subseteq \mathbb{C}$ that contains

an accumulation point, then $f(z)$ is identically zero. Also, we deduce that an analytic function is uniquely determined by its values on any set with an accumulation point in the region of analyticity.

Theorem B.5.7 (Conformal Mapping Theorem)**.** *A complex function is analytic if and only if it maps pairs of intersecting curves into pairs that intersect at the same angle.*

B.6. Quotients of the complex plane

In the theory of elliptic curves over $\mathbb{C}$, we often work with a quotient of the complex plane $\mathbb{C}$ modulo some lattice L. See Section 3.1 for the definition of lattice, the definition of the quotient $\mathbb{C}/L$ and the relationship to elliptic curves. In this section, we define what it means for a map $\mathbb{C}/L \to \mathbb{C}$ to be analytic.

Let $L \subset \mathbb{C}$ be a lattice with a basis $L = \langle w_1, w_2 \rangle$. Usually, we fix a fundamental domain for L as follows

$$\mathcal{F}_L = \{\lambda w_1 + \mu w_2 \in \mathbb{C} : 0 \leq \lambda, \mu < 1\}.$$

For our purposes here, we will define a fundamental domain for $\mathbb{C}/L$ for each $\alpha \in \mathbb{C}$ such that α is positioned in the interior of the domain:

$$\mathcal{F}_{L,\alpha} = \{\alpha + \lambda w_1 + \mu w_2 \in \mathbb{C} : -1/2 \leq \lambda, \mu < 1/2\}$$

and we also define the interior of $\mathcal{F}_{L,\alpha}$ by

$$\mathcal{F}^0_{L,\alpha} = \{\alpha + \lambda w_1 + \mu w_2 \in \mathbb{C} : -1/2 < \lambda, \mu < 1/2\}.$$

Notice that $\mathcal{F}^0_{L,\alpha}$ is a region in $\mathbb{C}$ (it is non-empty, connected and open), and α is at the center of the region. Notice that there is a bijection

$$\psi_{L,\alpha} : \mathbb{C}/L \to \mathcal{F}_{L,\alpha}. \tag{B.3}$$

Let $f : \mathbb{C}/L \to \mathbb{C}$ be a complex-valued function that is well-defined for every element of the quotient $\mathbb{C}/L$. Let $\alpha \bmod L$ be such an element. We say that $f : \mathbb{C}/L \to \mathbb{C}$ *is analytic at* α if the map

$$\widehat{f} : \mathcal{F}^0_{L,\alpha} \to \mathbb{C}, \quad \widehat{f}(z) = f(z \bmod L)$$

is analytic at α.

When we discuss maps between elliptic curves (e.g., Proposition 3.1.6), we talk about analytic maps $f : \mathbb{C}/L \to \mathbb{C}/L'$, where L and

L' are lattices. What does "analytic" mean in this context? How do we define analyticity? It is simply a matter of choosing correct charts for each $\mathbb{C}/L$ and $\mathbb{C}/L'$, as we shall see next.

Let $f : \mathbb{C}/L \to \mathbb{C}/L'$ be a continuous map. Let $\alpha \in \mathbb{C}$ and suppose that $f(\alpha \bmod L) = \beta \bmod L'$. Let $\mathcal{F}^0_{L,\alpha}$ be the region about α defined above, and similarly define $\mathcal{F}^0_{L',\beta}$. Let $\epsilon > 0$ be small enough so that the disc $D_\epsilon(\beta)$ is completely contained in $\mathcal{F}^0_{L',\beta}$. Then, by continuity of f, there is a δ such that if $|z - \alpha| < \delta$, then $f(z \bmod L)$ is inside $D_\epsilon(\beta)$. Pick δ small enough so that $D_\delta(\alpha)$ is completely contained in $\mathcal{F}^0_{L,\alpha}$. We are now ready to state our definition: we say that the continuous map $f : \mathbb{C}/L \to \mathbb{C}/L'$ *is analytic at* $\alpha \bmod L$ if the map

$$\widehat{f} : D_\delta(\alpha) \to D_\epsilon(\beta), \quad \widehat{f}(z) = \psi_{L',\beta}(f(z \bmod L)) \in D_\epsilon(\beta) \subseteq \mathcal{F}_{L',\beta}$$

is analytic at α, where $\psi_{L',\beta} : \mathbb{C}/L' \to \mathcal{F}_{L',\beta}$ is the bijection we defined in Eq. (B.3).

B.7. Exercises

Exercise B.7.1. The goal of this exercise is to prove that $\mathbb{C}$ is a field.

(1) Show that any non-zero complex number $\alpha = a + bi$ has a multiplicative inverse which is also a complex number $\alpha^{-1} = c + di$ with $c, d \in \mathbb{R}$.

(2) Convince yourself that $\mathbb{C}$ is a field; i.e., justify why $\mathbb{C}$ satisfies each of the field axioms.

Exercise B.7.2. Let α be a complex number. Show that $\alpha \in \mathbb{R}$ if and only if $\alpha = \overline{\alpha}$.

Exercise B.7.3. Show that $f(z) = z$ is analytic on $\mathbb{C}$. Also, show that $g_n(z) = z^n$ is analytic on $\mathbb{C}$ for every $n \geq 1$ and that the derivative is $g_n'(z) = nz^{n-1}$.

Exercise B.7.4. Let $f(z)$ be a complex-valued function that is analytic in a region $U \subseteq \mathbb{C}$.

(1) Show that f is also continuous at every point of U (i.e., $\lim_{h \to a} f(h) = f(a)$ for every $a \in U$).

(2) Let $f(z) = u(z) + v(z)i$, where $u(z)$ and $v(z)$ are real-valued. Show that u and v are continuous on U.

Exercise B.7.5. Let $f(z)$ be an analytic function on a region U, and write $\Re(f(z)) = u(z)$, $\Im(f(z)) = v(z)$ for the real and imaginary parts of $f(z)$, respectively. We define the *Laplacians* of u and v by

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad \text{and} \quad \Delta v = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}.$$

Show that $\Delta u = \Delta v = 0$. (Hint: use the Cauchy-Riemann differential equations, i.e., Eq. B.2.)

Exercise B.7.6. Prove the Fundamental Theorem of Algebra B.1.1: if $P(z)$ is a non-constant polynomial, then there is a root of P in $\mathbb{C}$. (Hint: suppose that $P(z)$ has no roots in $\mathbb{C}$. Then $1/P(z)$ would be analytic. Now use Liouville's Theorem B.5.5.)

Appendix C

Projective space

C.1. The projective line

Let us begin with an example. Consider the function $f(x) = \frac{1}{x}$. We know from Calculus that f is continuous (and differentiable) on all of its domain (i.e., $\mathbb{R}$) except at $x = 0$. Would it be possible to extend the real line so that $f(x)$ is continuous everywhere? The answer is yes, it is possible, and the solution is to *glue* the "end" of the real line at ∞ with the other "end" at $-\infty$. We will describe the solution in detail below. Formally, we need the *projective line*, which is a line with points $\mathbb{R} \cup \{\infty\}$, i.e., a real line plus a single point at infinity that ties the line together (into a circle).

The formal definition of the projective line is as follows. It may seem a little confusing at first, but it is fairly easy to work and compute with it. First, we need to define a relation between vectors of real numbers in the plane. Let a, b, x, y be real numbers such that neither (x, y) nor (a, b) is the zero vector. We say that $(x, y) \sim (a, b)$ if the vector (x, y) is a non-zero multiple of the vector (a, b). In other words, if we consider (a, b) and (x, y) as points in the plane, we say that $(a, b) \sim (x, y)$ if they both lie in one line on the plane that passes through the origin. Again:

$(x, y) \sim (a, b)$ if and only if there is $\lambda \in \mathbb{R}$ such that $x = \lambda a$, $y = \lambda b$.

For instance, $(\sqrt{2}, \sqrt{2}) \sim (1, 1)$. We denote by $[x, y]$ the set of all vectors (a, b) such that $(x, y) \sim (a, b)$:

$$[x, y] = \{(a, b) : a, b \in \mathbb{R} \text{ such that } (a, b) \neq (0, 0) \text{ and } (x, y) \sim (a, b)\}.$$

Finally, we define the real projective line by

$$\mathbb{P}^1(\mathbb{R}) = \{[x, y] : x, y \in \mathbb{R} \text{ with } (x, y) \neq (0, 0)\}.$$

If you think about it, $\mathbb{P}^1(\mathbb{R})$ is the set of all lines through the origin (each class $[x, y]$ consists of all points — except the origin — on the line that goes through (x, y) and $(0, 0)$). The important thing to notice is that if $[x, y] \in \mathbb{P}^1(\mathbb{R})$ and $y \neq 0$, then $(x, y) \sim (\frac{x}{y}, 1)$, so the class of $[x, y]$ contains a unique representative of the form $(a, 1)$ for some $a = \frac{x}{y} \in \mathbb{R}$. This allows the following decomposition of $\mathbb{P}^1(\mathbb{R})$:

$$\mathbb{P}^1(\mathbb{R}) = \{[x, 1] : x \in \mathbb{R}\} \cup \{[1, 0]\}.$$

The set of points $\{[x, 1]\}$ are in bijection with $\mathbb{R}$ and, therefore, form a real line. The point $[1, 0]$, which is the only point in $\mathbb{P}^1(\mathbb{R})$ that does not belong to the real line $\{[x, 1]\}$, is called the *point at infinity* (see Figure 1).

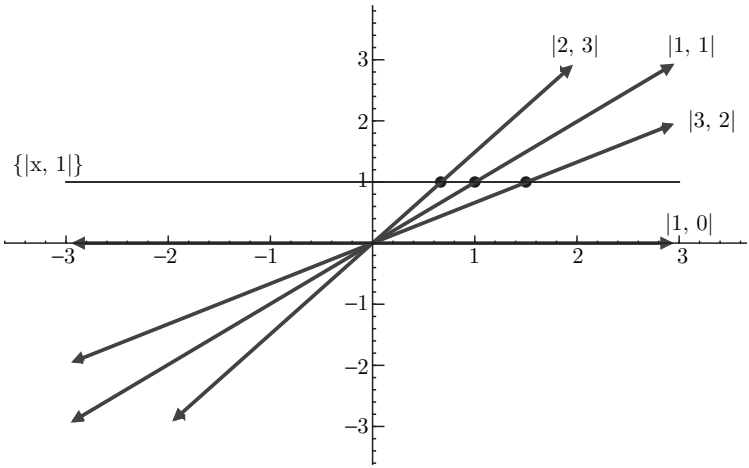


Figure 1. Some points in the projective line, e.g., $[2, 3] \in \mathbb{P}^1(\mathbb{R})$, and their representatives of the form $[x, 1]$, e.g. $[\frac{2}{3}, 1]$, except for $[1, 0]$.

Notice that when $x \in \mathbb{R}$ gets large (i.e., $x \to \infty$ or $x \to -\infty$), the point $[x, 1] \in \mathbb{P}^1(\mathbb{R})$ corresponds to a line in the real plane that is closer and closer to the horizontal line. Since the horizontal line corresponds to the point $[1, 0] \in \mathbb{P}^1(\mathbb{R})$, we see that as x gets large (in either the positive or negative direction!), the points $[x, 1]$ get closer and closer to $[1, 0]$, the point at infinity. This is what we meant at the beginning of this section by "glueing" both ends of the real line, ∞ and $-\infty$, at one point.

Let us see that, with this definition, the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 1/x$ is continuous everywhere when extended to $\mathbb{P}^1(\mathbb{R})$. We define instead an extended function $F : \mathbb{P}^1(\mathbb{R}) \to \mathbb{P}^1(\mathbb{R})$ by

$$F([x, y]) = [y, x].$$

Notice that a point on the real line of $\mathbb{P}^1$, i.e., a point of the form $[x, 1]$, is sent to the point $[1, x]$ of $\mathbb{P}^1$, and $(1, x) \sim (\frac{1}{x}, 1)$ as long as $x \neq 0$. So $[x, 1]$ with $x \neq 0$ is sent to $[\frac{1}{x}, 1]$ via F (i.e., the real point x is sent to $\frac{1}{x}$). Hence, F coincides with f on $\mathbb{R} - \{0\}$. But F is perfectly well-defined on $x = 0$, i.e., on the point $[0, 1]$, and $F([0, 1]) = [1, 0]$ so that $[0, 1]$ is sent to the point at infinity. Moreover, both sided limits coincide:

$$\lim_{x \to 0^+} F([x, 1]) = \lim_{x \to 0^-} F([x, 1]) = F([0, 1]) = [1, 0].$$

C.2. The projective plane

We may generalize the construction above of the projective line in order to construct a projective plane that will consist of a real plane plus a number of points at infinity, one for each direction in the plane; i.e., the projective plane will be a real plane plus a projective line of points at infinity.

Let $a, b, c, x, y, z \in \mathbb{R}$ such that neither (a, b, c) nor (x, y, z) are the zero vector:

$(x, y, z) \sim (a, b, c)$ if and only if there is $\lambda \in \mathbb{R}$ such that $x = \lambda a$, $y = \lambda b$, $z = \lambda c$.

We also define classes of similar vectors by

$[x, y, z] = \{(a, b, c) : a, b, c \in \mathbb{R}$ such that $(a, b, c) \neq \vec{0}$ and $(x, y, z) \sim (a, b, c)\}$.

Notice that, as before, the class $[x, y, z]$ contains all the points in the line in $\mathbb{R}^3$ that goes through (x, y, z) and $(0, 0, 0)$ except the origin. We define the projective plane to be the collection of all such lines:

$$\mathbb{P}^2(\mathbb{R}) = \{[x, y, z] : x, y, z \in \mathbb{R} \text{ such that } (x, y, z) \neq (0, 0, 0)\}.$$

If $z \neq 0$, then $(x, y, z) \sim (\frac{x}{z}, \frac{y}{z}, 1)$. Thus,

$$\mathbb{P}^2(\mathbb{R}) = \{[x, y, 1] : x, y \in \mathbb{R}\} \cup \{[a, b, 0] : a, b \in \mathbb{R}\}.$$

The points of the set $\{[x, y, 1] : x, y \in \mathbb{R}\}$ are in 1-to-1 correspondence with the real plane $\mathbb{R}^2$, and the points in $\{[a, b, 0] : a, b \in \mathbb{R}\}$ are called the points at infinity and form a $\mathbb{P}^1(\mathbb{R})$, a projective line.

One interesting consequence of the definitions is that any two parallel lines in the real plane $\{[x, y, 1]\}$ intersect at a point at infinity $[a, b, 0]$. Indeed, let $L : y = mx + b$ and $L' : y = mx + b'$ be distinct parallel lines in the real plane. If points in the real plane $\{[x, y, 1]\}$ correspond to lines in $\mathbb{R}^3$, then lines in the real plane correspond to *planes* in $\mathbb{R}^3$:

$$L = \{[x, y, z] : mx - y + bz = 0\}, \quad L' = \{[x, y, z] : mx - y + b'z = 0\}.$$

What is $L \cap L'$? The intersection points are those $[x, y, z]$ such that $mx - y + bz = mx - y + b'z = 0$, which implies that $(b - b')z = 0$. Since $L \neq L'$, we have $b \neq b'$ and, therefore, we must have $z = 0$. Hence

$$L \cap L' = \{[x, mx, 0] : x \in \mathbb{R}\} = \{[1, m, 0]\},$$

and so the intersection consists of a single point at infinity: $[1, m, 0]$.

C.3. Over an arbitrary field

The projective line and plane can be defined over any field. Let K be a field (e.g. $K = \mathbb{Q}, \mathbb{R}, \mathbb{C}$ or $\mathbb{F}_p$). The usual *affine plane* (or Euclidean plane) is defined by

$$\mathbb{A}^2(K) = \{(x, y) : x, y \in K\}.$$

The projective plane over K is defined by

$$\mathbb{P}^2(K) = \{[x, y, z] : x, y, z \in K \text{ such that } (x, y, z) \neq (0, 0, 0)\}.$$

As before, $(x, y, z) \sim (a, b, c)$ if and only if there is $\lambda \in K$ such that $(x, y, z) = \lambda \cdot (a, b, c)$.

C.4. Curves in the projective plane

Let K be a field and let C be a curve in affine space, given by a polynomial in two variables:

$$C : f(x, y) = 0$$

for some $f(x, y) \in K[x, y]$, e.g. $C : y^2 - x^3 - 1 = 0$. We want to extend C to a curve in the projective plane $\mathbb{P}^2(K)$. In order to do this, we consider the points in the curve (x, y) to be points in the plane $[\frac{x}{z}, \frac{y}{z}, 1]$ of $\mathbb{P}^2(K)$. Thus, we have

$$C : \left(\frac{y}{z}\right)^2 - \left(\frac{x}{z}\right)^3 - 1 = 0$$

or, equivalently, $zy^2 - x^3 - z^3 = 0$. Notice that the polynomial $F(x, y, z) = zy^2 - x^3 - z^3$ is homogeneous in its variables (each monomial has degree 3) and $F(x, y, 1) = f(x, y)$. The curve in $\mathbb{P}^2(K)$, given by

$$\widehat{C} : F(x, y, z) = zy^2 - x^3 - z^3 = 0,$$

is the curve we were looking for, which extends our original curve C in the affine plane. Notice that if the points $(x, y) \in C$, then $[x, y, 1] \in \widehat{C}$. However, there may be some extra points in $\widehat{C}$ which were not present in C, namely those points of $\widehat{C}$ at infinity. Recall that the points at infinity are those with $z = 0$, so $F(x, y, 0) = -x^3 = 0$ implies that $x = 0$ also, and the only point at infinity in $\widehat{C}$ is $[0, 1, 0]$.

In general, if $C \subseteq \mathbb{A}^2(K)$ is given by $f(x, y) = 0$ and d is the highest degree of a monomial in f, then $\widehat{C} \in \mathbb{P}^2(K)$ is given by

$$\widehat{C} : F(x, y, z) = 0,$$

where $F(x, y, z) = z^d \cdot f\left(\frac{x}{z}, \frac{y}{z}\right)$. Conversely, if $\widehat{C} : F(x, y, z) = 0$ is a curve in the projective plane, then $C : F(x, y, 1) = 0$ is a curve in the affine plane. In this case, C is the projection of $\widehat{C}$ onto the chart $z = 1$; we may also look at other charts, e.g., $x = 1$, which would yield a curve $C' : F(1, y, z) = 0$.

Here is another example. Let C be given by

$$C : y - x^2 = 0$$

so that C is a parabola. Then $\widehat{C}$ is given by

$$\widehat{C} : F(x, y, z) = z^2 f\left(\frac{x}{z}, \frac{y}{z}\right) = zy - x^2 = 0.$$

The curve $\widehat{C}$ has a unique point at infinity, namely $[0, 1, 0]$. This means that the two "arms" of the parabola meet at a single point at infinity. Thus, a parabola has the shape of an ellipse in $\mathbb{P}^2(K)$. How about hyperbolas? Let

$$C : x^2 - y^2 = 1.$$

Then $\widehat{C} : x^2 - y^2 = z^2$ and there are two points at infinity, namely $[1, 1, 0]$ and $[1, -1, 0]$. Thus, the four arms of the hyperbola in the affine plane meet in two points, and the hyperbola also has the shape of an ellipse in the projective plane $\mathbb{P}^2(K)$.

C.5. Singular and smooth curves

We say that a projective curve $C : F(x, y, z) = 0$ is singular at a point $P \in C$ if and only if $\frac{\partial F}{\partial x}(P) = \frac{\partial F}{\partial y}(P) = \frac{\partial F}{\partial z}(P) = 0$. In other words, C is singular at P if the tangent vector at P vanishes. Otherwise, we say that C is non-singular at P. If C is non-singular at every point, we say that C is a smooth (or non-singular) curve.

For example, $C : zy^2 = x^3$ is singular at $P = [0, 0, 1]$ because $F(x, y, z) = zy^2 - x^3$ and

$$\frac{\partial F}{\partial x} = -x^2, \ \frac{\partial F}{\partial y} = 2yz, \ \frac{\partial F}{\partial z} = y^2.$$

Thus, $\frac{\partial F}{\partial x}(P) = \frac{\partial F}{\partial y}(P) = \frac{\partial F}{\partial z}(P) = 0$ for $P = [0, 0, 1]$.

Here is another example. The curve $D : z^2y^2 = x^4 + z^4$ has partial derivatives

$$\frac{\partial F}{\partial x} = -4x^3, \ \frac{\partial F}{\partial y} = 2yz^2, \ \frac{\partial F}{\partial z} = 2y^2z - 4z^3.$$

Thus, if $P = [x, y, z] \in D(\mathbb{Q})$ is singular, then

$$-4x^3 = 0, \quad 2yz^2 = 0, \quad \text{and} \quad 2y^2z - 4z^3 = 0.$$

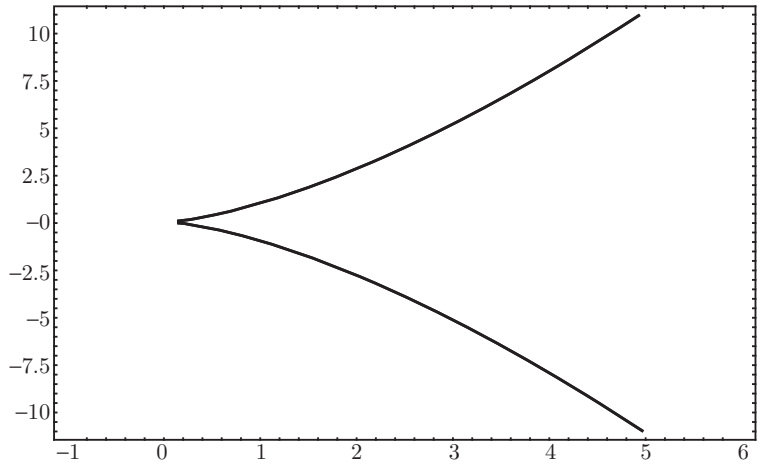


Figure 2. The chart $\{[x, y, 1]\}$ of the curve $zy^2 = x^3$.

The first two equalities imply that $x = 0$ and $yz = 0$ (what would happen if we were working over a field of characteristic 2, such as $\mathbb{F}_2$?). If $y = 0$, then $z = 0$ by the third equation, but $[0, 0, 0]$ is not a well-defined point in $\mathbb{P}^2(\mathbb{Q})$, so this is impossible. However, if $x = z = 0$, then y may take any value. Hence, $P = [0, 1, 0]$ is a singular point. Notice that the affine curve that corresponds to the chart $z = 1$ of D, given by $y^2 = x^4 + 1$, is non-singular at all points in the affine plane but is singular at a point at infinity, namely $P = [0, 1, 0]$.

An elliptic curve of the form $E : y^2 = x^3 + Ax + B$, or in projective coordinates given by $zy^2 = x^3 + Axz^2 + Bz^3$, is non-singular if and only if $4A^3 + 27B^2 \neq 0$. The quantity $\Delta = -16 \cdot (4A^3 + 27B^2)$ is called the discriminant of E.

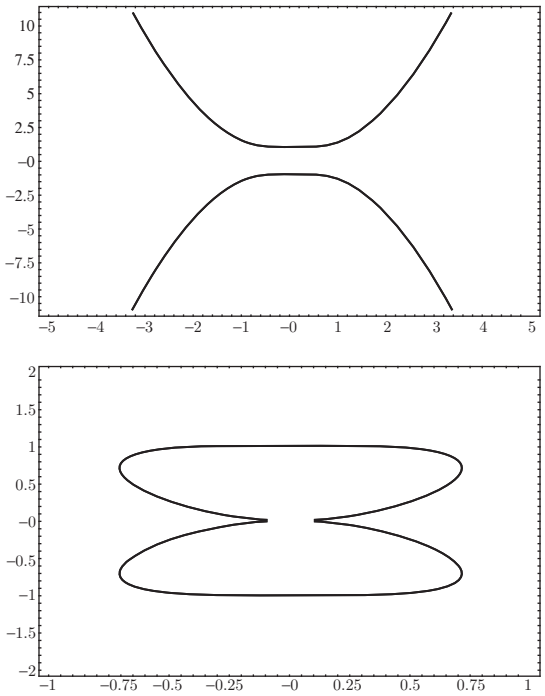


Figure 3. The chart $\{[x, y, 1]\}$ of the curve $z^2y^2 = x^4 + z^4$ (above, non-singular) and the chart $\{[x, 1, z]\}$ (below, the curve is singular).

Appendix D

The p-adic numbers

In this appendix we briefly introduce the p-adic integers $\mathbb{Z}_p$ and the p-adic numbers $\mathbb{Q}_p$. We strongly recommend [**Gou97**] to learn more about the p-adics.

Let $p \geq 2$ be a prime. The p-adic numbers may be thought of as a generalization of $\mathbb{Z}/p\mathbb{Z}$. The main difference is that the p-adic numbers form a ring of characteristic zero, while $\mathbb{Z}/p\mathbb{Z}$ has characteristic p. In $\mathbb{Z}/p\mathbb{Z}$ we only consider congruences modulo p, while in $\mathbb{Z}_p$ we consider congruences modulo p^n for all $n > 0$. The p-adic integers, denoted by $\mathbb{Z}_p$, are defined as follows:

$$\mathbb{Z}_p = \{(a_1, a_2, \ldots) : a_n \in \mathbb{Z}/p^n\mathbb{Z} \text{ such that } a_{n+1} \equiv a_n \bmod p^n\}.$$

In other words, a p-adic integer is an infinite vector $(a_n)_{n=1}^{\infty}$ such that the nth coordinate belongs to $\mathbb{Z}/p^n\mathbb{Z}$ and the sequence is coherent under congruences; i.e., $a_{n+1} \in \mathbb{Z}/p^{n+1}\mathbb{Z}$ reduces to the previous term a_n modulo p^n. For instance,

$$(2, 2, 29, 29, 272, 758, \ldots)$$

are the first few terms of a 3-adic integer; notice that all the coordinates are coherent with the previous terms under congruences modulo powers of 3. The vector $(2, 2, 2, 2, \ldots)$ is another element of $\mathbb{Z}_3$ (which we will denote simply by 2).

The p-adic integers have addition and multiplication operations, defined coordinate-by-coordinate:

$$(a_n)_{n=1}^{\infty} + (b_n)_{n=1}^{\infty} = (a_n + b_n \bmod p^n)_{n=1}^{\infty},$$

and

$$(a_n)_{n=1}^{\infty} \cdot (b_n)_{n=1}^{\infty} = (a_n \cdot b_n \bmod p^n)_{n=1}^{\infty}.$$

The reader should check that the sum and product of two coherent vectors is also coherent under congruences and, therefore, a new element of $\mathbb{Z}_p$. These operations make $\mathbb{Z}_p$ a commutative ring with identity element $1 = (1, 1, 1, 1, \ldots)$ and zero element $0 = (0, 0, 0, 0, \ldots)$.

For any prime $p \geq 2$, the p-adic integers contain a copy of $\mathbb{Z}$, where the integer m is represented by the element

$$m = (m \bmod p,\ m \bmod p^2,\ m \bmod p^3, \ldots).$$

For example, the number 200 in $\mathbb{Z}_3$ is given by

$$200 = (2, 2, 11, 38, 200, 200, 200, 200, 200, 200, \ldots).$$

Thus, we may write $\mathbb{Z} \subseteq \mathbb{Z}_p$ (see Exercise D.2.1). However, there are elements in $\mathbb{Z}_p$ that are not in $\mathbb{Z}$, so $\mathbb{Z} \subsetneq \mathbb{Z}_p$. Here is an example for $p = 7$: we are going to show that $\mathbb{Z}_7$, unlike $\mathbb{Z}$, contains an element whose square is 2 (which we will denote by "$\sqrt{2}$"). Indeed, 2 is a quadratic residue in $\mathbb{Z}/7\mathbb{Z}$, and 2 has two square roots, namely 3 and 4 modulo 7. A standard theorem of number theory shows that, hence, 2 is in fact a quadratic residue modulo 7^n for all $n \geq 1$. Thus, there exist integers a_n such that $a_n^2 \equiv 2 \bmod p^n$ for all $n \geq 1$. Moreover, it can also be shown that, if a_n is chosen, then there is $a_{n+1} \bmod p^{n+1}$ with $a_{n+1}^2 \equiv 2 \bmod p^{n+1}$ and $a_{n+1} \equiv a_n \bmod p^n$ (we say that a_n can be *lifted* to $\mathbb{Z}/p^{n+1}\mathbb{Z}$; see Exercise D.2.2). Indeed, here are the first few coordinates of an element α of $\mathbb{Z}_7$ such that $\alpha^2 = (2, 2, 2, \ldots)$:

$$\alpha = (3, 10, 108, 2166, 4567, \ldots).$$

Thus, α should be regarded as "$\sqrt{2}$ " inside $\mathbb{Z}_7$, and $-\alpha$ is another square root of 2.

The usual integers, $\mathbb{Z}$, are not a field because not every element has a multiplicative inverse (only ± 1 have inverses!). Similarly, the p-adic integers $\mathbb{Z}_p$ do not form a field either; e.g., $p = (p, p, p, \ldots)$ is not invertible in $\mathbb{Z}_p$, but many elements of $\mathbb{Z}_p$ *are* invertible. For

instance, if $p > 2$, then 2 is invertible in $\mathbb{Z}_p$ (in other words, there is a number $\frac{1}{2} \in \mathbb{Z}_p$). Indeed, the inverse of 2 is given by

$$\frac{1}{2} = \left(\frac{1+p}{2}, \frac{1+p^2}{2}, \ldots, \frac{1+p^n}{2}, \ldots \right).$$

For example, in $\mathbb{Z}_5$, the inverse of 2 is given by $(3, 13, 63, 313, \ldots)$. It is easy to see that if $\alpha = (a_n)_{n=1}^{\infty}$ with $a_1 \not\equiv 0 \bmod p$, then α is invertible in $\mathbb{Z}_p$. If $a_1 \equiv 0 \bmod p$, then α is not invertible. Moreover, for any $\alpha \in \mathbb{Z}_p$ there is an $r \geq 0$ such that $\alpha = p^r \beta$, where $\beta \in \mathbb{Z}_p$ is invertible.

Even though $\mathbb{Z}_p$ is not a field, we can embed $\mathbb{Z}_p$ in a field in the same way that $\mathbb{Z}$ sits inside $\mathbb{Q}$. We define the field of p-adic numbers by

$$\mathbb{Q}_p = \left\{ \frac{\alpha}{p^k} : k \geq 0 \text{ and } \alpha \in \mathbb{Z}_p \right\}.$$

Thus, every element of $\alpha \in \mathbb{Q}_p$ can be written as $\alpha = p^r \beta$ with $r \in \mathbb{Z}$ and an invertible $\beta \in \mathbb{Z}_p^{\times}$.

D.1. Hensel's lemma

The following results are used to show the existence of a solution to polynomial equations over *local fields*. Here we will only discuss the application to the p-adics, $\mathbb{Q}_p$ (which is an example of a local field). Notice the similarities with Newton's method.

Theorem D.1.1 (Hensel's Lemma)**.** *Let $p \geq 2$, let $\mathbb{Q}_p$ be the field of p-adic numbers and let $\mathbb{Z}_p$ be the p-adic integers. Let ν_p be the usual p-adic valuation (i.e., $\nu_p(p^e n) = e$ if $n \in \mathbb{Z}$ and $\gcd(n, p) = 1$). Let $f(x)$ be a polynomial with coefficients in $\mathbb{Z}_p$ and suppose there exist $\alpha_0 \in \mathbb{Z}_p$ such that*

$$\nu_p(f(\alpha_0)) > \nu_p(f'(\alpha_0)^2).$$

Then there exists a root $\alpha \in \mathbb{Q}_p$ of $f(x)$. Moreover, the sequence

$$\alpha_{i+1} = \alpha_i - \frac{f(\alpha_i)}{f'(\alpha_i)}$$

converges to α. Furthermore;

$$\nu_p(\alpha - \alpha_0) \geq \nu_p\left(\frac{f(\alpha_i)}{f'(\alpha_i)} \right) > 0.$$

Corollary D.1.2 (Trivial case of Hensel's lemma)**.** *Let $p \geq 2$, and let $\mathbb{Z}_p$ and $\mathbb{Q}_p$ be as before. Let $f(x)$ be a polynomial with coefficients in $\mathbb{Z}_p$ and suppose there exist $\alpha_0 \in \mathbb{Z}_p$ such that*

$$f(\alpha_0) \equiv 0 \ mod\ p, \quad f'(\alpha_0) \neq 0 \ mod\ p.$$

Then there exists a root $\alpha \in \mathbb{Q}_p$ of $f(x)$, i.e., $f(\alpha) = 0$.

Example D.1.3. Let p be a prime number greater than 2. Are there solutions to $x^2 + 7 = 0$ in the field $\mathbb{Q}_p$? If there are, -7 must be a quadratic residue modulo p. Thus, let p be a prime such that

$$\left(\frac{-7}{p}\right) = 1,$$

where $(\frac{\cdot}{p})$ is Legendre's quadratic residue symbol. Hence, there exist $\alpha_0 \in \mathbb{Z}$ such that $\alpha_0^2 \equiv -7 \bmod p$. We claim that $x^2 + 7 = 0$ has a solution in $\mathbb{Q}_p$ if and only if -7 is a quadratic residue modulo p. Indeed, if we let $f(x) = x^2 + 7$ (so $f'(x) = 2x$), the element $\alpha_0 \in \mathbb{Z}_p$ satisfies the conditions of the (trivial case of) Hensel's lemma. Therefore, there exists a root $\alpha \in \mathbb{Q}_p$ of $x^2 + 7 = 0$. ■

Example D.1.4. Let $p = 2$. Are there any solutions to $x^2 + 7 = 0$ in $\mathbb{Q}_2$? Notice that if we let $f(x) = x^2 + 7$, then $f'(x) = 2x$ and, for any $\alpha_0 \in \mathbb{Z}_2$, the number $f'(\alpha_0) = 2\alpha_0$ is congruent to 0 modulo 2. Thus, we cannot use the trivial case of Hensel's lemma (i.e., Corollary D.1.2).

Let $\alpha_0 = 1 \in \mathbb{Z}_2$. Notice that $f(1) = 8$ and $f'(1) = 2$. Thus,

$$3 = \nu_2(8) > \nu_2(2^2) = 2$$

and the general case of Hensel's lemma applies. Hence, there exists a 2-adic solution to $x^2 + 7 = 0$. ■

D.2. Exercises

Exercise D.2.1. Show that if q and t are distinct integers (in $\mathbb{Z}$), then their representatives in $\mathbb{Z}_p$ for any prime $p \geq 2$, given by $q = (q \bmod p^n)_{n=1}^{\infty}$ and $t = (t \bmod p^n)_{n=1}^{\infty}$, are also distinct in $\mathbb{Z}_p$.

Exercise D.2.2. Let $p > 2$ be a prime number.

(1) Let $b \in \mathbb{Z}$ with $\gcd(b, p) = 1$, and let $n \geq 1$. Suppose $a_n \in \mathbb{Z}$ such that $a_n^2 \equiv b \bmod p^n$. Show that there exists $a_{n+1} \in \mathbb{Z}$ such that $a_{n+1}^2 \equiv b \bmod p^{n+1}$ and $a_{n+1} \equiv a_n \bmod p^n$. (Hint: write $a_n^2 = b + kp^n$ and consider $f(x) = a_n + xp^n$. Find x such that $f(x)^2 \equiv b \bmod p^{n+1}$.)

(2) Suppose $a_1^2 \equiv b \bmod p$, where $\gcd(b, p) = 1$. Show that the vector $\alpha = (a_n)_{n=1}^{\infty}$, defined recursively by

$$a_{n+1} = a_n - \frac{a_n^2 - b}{2a_n} \bmod p^{n+1},$$

is a well-defined element of $\mathbb{Z}_p$ and, moreover, $\alpha^2 = b$, i.e.,

$$\alpha^2 = (b \bmod p,\ b \bmod p^2,\ b \bmod p^3, \ldots),$$

so α is a square root of b.

Exercise D.2.3. Find the first 4 coordinates of the 5-adic expansion of $\frac{1}{3}$ in $\mathbb{Z}_5$.

Exercise D.2.4. Find the first 4 coordinates of the 5-adic expansions of $\pm\sqrt{6}$ in $\mathbb{Z}_5$; i.e., find the first 4 coordinates of α and $-\alpha$ such that $\alpha^2 = 6$ in $\mathbb{Z}_5$.

Appendix E

Parametrization of torsion structures

In this appendix we provide one-parameter infinite families of elliptic curves with all the possible torsion subgroups that may occur for elliptic curves over $\mathbb{Q}$. The main reference for this appendix is [**Kub76**], Table 3, p. 217.

In each table below, Figure 1 and Figure 2, we provide elliptic curves $E_{a,b}$ whose equations depend on two rational parameters $a, b \in \mathbb{Q}$, and such that the torsion subgroup $E_{a,b}(\mathbb{Q})_{\text{tors}}$ has a given subgroup G; i.e., the full torsion subgroup contains G as a subgroup, but may be larger in certain cases (see Example E.1.1 below).

The families that appear in Figure 1 depend on two independent parameters a, b, and the only condition that needs to be satisfied is that the discriminant $\Delta_{a,b}$ of $E_{a,b}$ must be non-zero. This condition on the discriminant is given in the second column of the table.

The families that appear in Figure 2 depend on one single rational parameter $t \in \mathbb{Q}$, and a and b are rational functions in the variable t. The curves $E_{a,b}$ that appear in this table are all of the form

$$E_{a,b} : y^2 + (1-a)xy - by = x^3 - bx^2.$$

The point $(0, 0)$ is a torsion point of the maximal order in the group. The discriminant $\Delta_{a,b}$ of $E_{a,b}$ is always assumed to be non-zero.

$E_{a,b}/\mathbb{Q}$	$\Delta_{a,b} \neq 0$	G
$y^2 = x^3 + ax^2 + bx$	$a^2b^2 - 4b^3 \neq 0$	$\mathbb{Z}/2\mathbb{Z}$
$y^2 + axy + by = x^3$	$a^3b^3 - 27b^4 \neq 0$	$\mathbb{Z}/3\mathbb{Z}$
$y^2 = x(x+a)(x+b)$	$0 \neq a \neq b \neq 0$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$

Figure 1. Two-parameter families of elliptic curves $E_{a,b}/\mathbb{Q}$ such that $E_{a,b}(\mathbb{Q})_{\text{tors}}$ has a subgroup G.

Curves of the form $E_{a,b} : y^2 + (1-a)xy - by = x^3 - bx^2$		
a	b	G
$a = 0$	$b = t$	$\mathbb{Z}/4\mathbb{Z}$
$a = t$	$b = t$	$\mathbb{Z}/5\mathbb{Z}$
$a = t$	$b = t + t^2$	$\mathbb{Z}/6\mathbb{Z}$
$a = t^2 - t$	$b = t^3 - t^2$	$\mathbb{Z}/7\mathbb{Z}$
$a = \frac{(2t-1)(t-1)}{t}$	$b = (2t-1)(t-1)$	$\mathbb{Z}/8\mathbb{Z}$
$a = t^2(t-1)$	$b = t^2(t-1)(t^2-t+1)$	$\mathbb{Z}/9\mathbb{Z}$
$a = \frac{t(t-1)(2t-1)}{t^2-3t+1}$	$b = \frac{t^3(t-1)(2t-1)}{(t^2-3t+1)^2}$	$\mathbb{Z}/10\mathbb{Z}$
$a = \frac{t(1-2t)(3t^2-3t+1)}{(t-1)^3}$	$b = -a \cdot \frac{2t^2-2t+1}{t-1}$	$\mathbb{Z}/12\mathbb{Z}$
$a = 0$	$b = t^2 - \frac{1}{16}$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}$
$a = \frac{10-2t}{t^2-9}$	$b = \frac{-2(t-1)^2(t-5)}{(t^2-9)^2}$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z}$
$a = \frac{(2t+1)(8t^2+4t+1)}{2(4t+1)(8t^2-1)t}$	$b = \frac{(2t+1)(8t^2+4t+1)}{(8t^2-1)^2}$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/8\mathbb{Z}$

Figure 2. One-parameter families of elliptic curves $E_{a,b}/\mathbb{Q}$ such that $E_{a,b}(\mathbb{Q})_{\text{tors}}$ has a subgroup G.

Example E.1.1. For each $t \in \mathbb{Q}$, according to Figure 2, the torsion subgroup of the elliptic curve $E_{t,t} : y^2 + (1-t)xy - ty = x^3 - tx^2$ contains $G = \mathbb{Z}/5\mathbb{Z}$ as a subgroup, as long as the discriminant $\Delta_{t,t} = t^5(t^2 - 11t - 1)$ is non-zero (thus, $\Delta_{t,t} = 0$ if and only if $t = 0$). In other words, the point $(0, 0)$ of $E_{t,t}$ is a torsion point of order 5.

Notice, however, that this does not imply that the torsion subgroup of $E_{t,t}(\mathbb{Q})$ is identical to $\mathbb{Z}/5\mathbb{Z}$. For instance, let $t = 12$. The torsion subgroup of the elliptic curve

$$E_{12,12} : y^2 - 11xy - 12y = x^3 - 12x^2$$

is isomorphic to $\mathbb{Z}/10\mathbb{Z}$. The point $(0, 0)$ is a point of order 5, but the point $(-6, -18)$ has exact order 10.

Example E.1.2. According to Figure 2, each curve in the family

$$y^2 + (1 + t - t^2)xy + (t^2 - t^3)y = x^3 + (t^2 - t^3)x^2$$

has a torsion point of exact order 7, namely $P = (0, 0)$, as long as the discriminant $\Delta = t^7(t - 1)^7(t^3 - 8t^2 + 5t + 1)$ is non-zero, which can only happen for the rational values $t = 0$ and $t = 1$. By Mazur's Theorem 2.5.2, the only possible torsion subgroup for an elliptic curve over $\mathbb{Q}$ that contains $\mathbb{Z}/7\mathbb{Z}$ as a subgroup is $\mathbb{Z}/7\mathbb{Z}$ itself. Thus, the torsion subgroup of each elliptic curve in this family is exactly $\mathbb{Z}/7\mathbb{Z}$.

Similarly, if $E_{a,b}$ is an elliptic curve in one of the families in Figure 2 that correspond to G in the list

$$\mathbb{Z}/7\mathbb{Z},\ \mathbb{Z}/9\mathbb{Z},\ \mathbb{Z}/10\mathbb{Z},\ \mathbb{Z}/12\mathbb{Z},\ \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z},\ \text{ or }\ \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/8\mathbb{Z},$$

then the torsion subgroup of $E_{a,b}(\mathbb{Q})$ must be exactly G.

Bibliography

[Ahl79] Lars Ahlfors, *Complex Analysis*, McGraw-Hill Science/Engineering/Math; 3rd edition (1979).

[AL70] A. O. L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. 185 (1970), 134-160.

[Bak90] Alan Baker, *Transcendental Number Theory*, Cambridge University Press, Cambridge, 1990.

[BSD63] B. Birch and H. P. F. Swinnerton-Dyer, *Notes on elliptic curves (I) and (II)*, J. Reine Angew. Math. 212 (1963), pp. 7-25, and 218 (1965), pp. 79-108.

[BCDT01] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor, *On the modularity of elliptic curves over Q: Wild 3-adic exercises*, Journal of the American Mathematical Society 14 (2001), pp. 843-939.

[Cha06] Jasbir S. Chahal, *Congruent numbers and elliptic curves*, Amer. Math. Monthly 113 (2006), no. 4, pp. 308-317.

[Chi95] Lindsay N. Childs, *A Concrete Introduction to Higher Algebra*, Springer, New York, 1995.

[Con08] Keith Conrad, *The congruent number problem*, available at http://www.math.uconn.edu/∼kconrad/blurbs/ugradnumthy/congnumber.pdf

[CSS00] Gary Cornell, Joseph H. Silverman and Glenn Stevens (Editors), *Modular Forms and Fermat's Last Theorem*, Springer, 2000.

[Cre97] John Cremona, *Algorithms for Modular Elliptic Curves*, Cambridge University Press, 1997 (available for free online).

[DS05] Fred Diamond and Jerry Shurman, *A First Course in Modular Forms*, Springer, New York, 2005.

[Dic05] Leonard E. Dickson, *History of the Theory of Numbers, Volume II: Diophantine Analysis*, Dover Publications, 2005.

[Dok04] Tim Dokchitser, *Computing special values of motivic L-functions*, Exper. Math. 13, No. 2 (2004), 137-149.

[Dou98] Darrin Doud, *A procedure to calculate torsion of elliptic curves over* $\mathbb{Q}$, Manuscripta Math. 95 (1998), 463-469.

[Duj09] Andrej Dujella's website, at http://web.math.hr/∼duje/tors/tors.html

[Fre86] Gerhard Frey, *Links between solutions of* $A-B=C$ *and elliptic curves*. Number theory (Ulm, 1987), 31–62, Lecture Notes in Math., 1380, Springer, New York, 1989.

[Gou97] Fernando Gouvea, *p-adic Numbers: An Introduction*, Springer (Universitext), 2nd edition (1997).

[GJPST09] Grigor Grigorov, Andrei Jorza, Stefan Patrikis, William A. Stein, and Corina Tarnita, *Computational verification of the Birch and Swinnerton-Dyer conjecture for individual elliptic curves*, Math. Comp. 78 (2009), 2397-2425.

[Kob93] Neal I. Koblitz, *Introduction to Elliptic Curves and Modular Forms*, Second Edition, Springer-Verlag, New York, 1993.

[Kub76] Daniel S. Kubert, *Universal bounds on the torsion of elliptic curves*, Proc. London Math. Soc. (3), 33, 1976, p. 193-237.

[Lan83] Serge Lang, *Conjectured Diophantine estimates on elliptic curves*, Progress in Math. 35, Birkhäuser, 1983.

[Li75] Wen-Ching Winnie Li, *Newforms and functional equations*, Math. Ann. 212 (1975), 285-315.

[Loz05] Álvaro Lozano-Robledo, *Buscando puntos racionales en curvas elípticas: Métodos explícitos*, La Gaceta de la Real Sociedad Matematica Española (J. of the Royal Mathematical Society of Spain), Vol. 8 (2005), n. 2, pp. 471-488.

[Loz08] Julian Aguirre, Álvaro Lozano-Robledo and Juan Carlos Peral, *Elliptic curves of maximal rank*, in Revista Matemática Iberoamericana, proceedings of the conference "Segundas Jornadas de Teoria de Números".

[Lut37] E. Lutz, *Sur l'equation* $y^2 = x^3 - Ax - B$ *dans les corps p-adic*, J. Reine Angew. Math. 177 (1937), 431-466.

[Mat93] Yuri V. Matiyasevich, *Hilbert's Tenth Problem*, MIT Press, Cambridge, Massachusetts, 1993.

[Maz72] Barry Mazur, *Courbes elliptiques et symboles modulaires*, Lecture Notes in Mathematics, Vol. 317, 277-294.

[Maz77] Barry Mazur, *Modular curves and the Eisenstein ideal*, IHES Publ. Math. 47 (1977), 33-186.

[Maz78] Barry Mazur, *Rational isogenies of prime degree*, Invent. Math. 44 (1978), 129-162.

[Mil10] Robert L. Miller, *Empirical Evidence for the Birch and Swinnerton-Dyer Conjecture*, Ph.D. Thesis, 2010.

[Mil06] J. S. Milne, *Elliptic Curves*, Kea Books, 2006.

[Miy06] T. Miyake, *Modular Forms*, Springer Monographs in Mathematics, New York, 2006.

[Nag35] T. Nagell, *Solution de quelque problemes dans la theorie arithmetique des cubiques planes du premier genre*, Wid. Akad. Skrifter Oslo I, 1935, Nr. 1.

[Rib90] Kenneth A. Ribet, *On modular representations of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ *arising from modular forms*. Invent. Math. 100 (1990), no. 2, 431–476.

[RuS02] Karl Rubin and Alice Silverberg, *Ranks of Elliptic Curves*, Bull. Amer. Math. Soc. 39, no. 4, pg. 455-474.

[SK52] J. G. Semple and G. T. Kneebone, *Algebraic Projective Geometry*, Oxford University Press, USA (1952).

[Ser77] J.-P. Serre, *A course in arithmetic*, Springer-Verlag, New York, 1973.

[Ser87] J.-P. Serre, *Sur les représentations modulaires de degré* 2 *de* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. (French) [On modular representations of degree 2 of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$] Duke Math. J. 54 (1987), no. 1, 179–230.

[Ser97] J.-P. Serre, *Galois Cohomology*, Springer-Verlag, New York, 1997.

[ShT67] I. R. Shafarevich and J. Tate, *The rank of elliptic curves*, AMS Transl. 8 (1967), 917-920.

[Shi73] Goro Shimura, *Introduction to Arithmetic Theory of Automorphic Functions*, Princeton University Press, 1973.

[Shi02] Goro Shimura, *The Representation of Integers as Sums of Squares*, American Journal of Mathematics, Vol. 124, No. 5 (Oct., 2002), pp. 1059-1081.

[Sil86] Joseph H. Silverman, *The Arithmetic of Elliptic Curves*, Springer-Verlag, New York, 1986.

[Sil94] Joseph H. Silverman, *Advanced Topics in the Arithmetic of Elliptic Curves*, Springer-Verlag, New York, 1994.

[SiT92] Joseph H. Silverman and John Tate, *Rational Points on Elliptic Curves*, Springer-Verlag, New York, 1992.

[Ste07] W. Stein, *Modular Forms, a computational approach*, American Mathematical Society, 2007.

[Ste08] W. Stein, *Elementary Number Theory: Primes, Congruences, and Secrets: A Computational Approach*, Undergraduate Texts in Mathematics, Springer, New York, 2008.

[Ste75] N. M. Stephens, *Congruence properties of congruent numbers*, Bull. London Math. Soc. 7 (1975), pp. 182-184.

[Tat74] J. Tate, *The arithmetic of elliptic curves*, Invent. Math., 23 (1974), 179-206.

[TW95] Richard Taylor and Andrew Wiles, *Ring-theoretic properties of certain Hecke algebras.* Ann. of Math. (2) 141 (1995), no. 3, 553-572.

[Tun83] J. Tunnell, *A Classical Diophantine Problem and Modular Forms of Weight* 3/2, Invent. Math. 72 (1983), pp. 323-334.

[Ver05] H. Verrill, *Fundamental Domain Drawer Applet* http://www.math.uconn.edu/∼alozano/fundomain/

[Was08] L. C. Washington, *Elliptic Curves: Number Theory and Cryptography*, Second Edition (Discrete Mathematics and Its Applications), Chapman & Hall/CRC (April 3, 2008).

[Wil95] Andrew Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. 141 (1995), no. 3, pp. 443-551.

Index

Titles in This Series

58 **Álvaro Lozano-Robledo,** Elliptic curves, modular forms, and their L-functions, 2011

57 **Charles M. Grinstead, William P. Peterson, and J. Laurie Snell,** Probability tales, 2011

56 **Julia Garibaldi, Alex Iosevich, and Steven Senger,** The Erdős distance problem, 2011

55 **Gregory F. Lawler,** Random walk and the heat equation, 2010

54 **Alex Kasman,** Glimpses of soliton theory: The algebra and geometry of nonlinear PDEs, 2010

53 **Jiří Matoušek,** Thirty-three miniatures: Mathematical and algorithmic applications of linear algebra, 2010

52 **Yakov Pesin and Vaughn Climenhaga,** Lectures on fractal geometry and dynamical systems, 2009

51 **Richard S. Palais and Robert A. Palais,** Differential equations, mechanics, and computation, 2009

50 **Mike Mesterton-Gibbons,** A primer on the calculus of variations and optimal control theory, 2009

49 **Francis Bonahon,** Low-dimensional geometry: From euclidean surfaces to hyperbolic knots, 2009

48 **John Franks,** A (terse) introduction to Lebesgue integration, 2009

47 **L. D. Faddeev and O. A. Yakubovskiĭ,** Lectures on quantum mechanics for mathematics students, 2009

46 **Anatole Katok and Vaughn Climenhaga,** Lectures on surfaces: (Almost) everything you wanted to know about them, 2008

45 **Harold M. Edwards,** Higher arithmetic: An algorithmic introduction to number theory, 2008

44 **Yitzhak Katznelson and Yonatan R. Katznelson,** A (terse) introduction to linear algebra, 2008

43 **Ilka Agricola and Thomas Friedrich,** Elementary geometry, 2008

42 **C. E. Silva,** Invitation to ergodic theory, 2007

41 **Gary L. Mullen and Carl Mummert,** Finite fields and applications, 2007

40 **Deguang Han, Keri Kornelson, David Larson, and Eric Weber,** Frames for undergraduates, 2007

39 **Alex Iosevich,** A view from the top: Analysis, combinatorics and number theory, 2007

38 **B. Fristedt, N. Jain, and N. Krylov,** Filtering and prediction: A primer, 2007

37 **Svetlana Katok,** p-adic analysis compared with real, 2007

36 **Mara D. Neusel,** Invariant theory, 2007

35 **Jörg Bewersdorff,** Galois theory for beginners: A historical perspective, 2006

TITLES IN THIS SERIES

34 **Bruce C. Berndt,** Number theory in the spirit of Ramanujan, 2006

33 **Rekha R. Thomas,** Lectures in geometric combinatorics, 2006

32 **Sheldon Katz,** Enumerative geometry and string theory, 2006

31 **John McCleary,** A first course in topology: Continuity and dimension, 2006

30 **Serge Tabachnikov,** Geometry and billiards, 2005

29 **Kristopher Tapp,** Matrix groups for undergraduates, 2005

28 **Emmanuel Lesigne,** Heads or tails: An introduction to limit theorems in probability, 2005

27 **Reinhard Illner, C. Sean Bohun, Samantha McCollum, and Thea van Roode,** Mathematical modelling: A case studies approach, 2005

26 **Robert Hardt, Editor,** Six themes on variation, 2004

25 **S. V. Duzhin and B. D. Chebotarevsky,** Transformation groups for beginners, 2004

24 **Bruce M. Landman and Aaron Robertson,** Ramsey theory on the integers, 2004

23 **S. K. Lando,** Lectures on generating functions, 2003

22 **Andreas Arvanitoyeorgos,** An introduction to Lie groups and the geometry of homogeneous spaces, 2003

21 **W. J. Kaczor and M. T. Nowak,** Problems in mathematical analysis III: Integration, 2003

20 **Klaus Hulek,** Elementary algebraic geometry, 2003

19 **A. Shen and N. K. Vereshchagin,** Computable functions, 2003

18 **V. V. Yaschenko, Editor,** Cryptography: An introduction, 2002

17 **A. Shen and N. K. Vereshchagin,** Basic set theory, 2002

16 **Wolfgang Kühnel,** Differential geometry: curves – surfaces – manifolds, second edition, 2006

15 **Gerd Fischer,** Plane algebraic curves, 2001

14 **V. A. Vassiliev,** Introduction to topology, 2001

13 **Frederick J. Almgren, Jr.,** Plateau's problem: An invitation to varifold geometry, 2001

12 **W. J. Kaczor and M. T. Nowak,** Problems in mathematical analysis II: Continuity and differentiation, 2001

11 **Mike Mesterton-Gibbons,** An introduction to game-theoretic modelling, 2000

10 **John Oprea,** The mathematics of soap films: Explorations with Maple®, 2000

9 **David E. Blair,** Inversion theory and conformal mapping, 2000

For a complete list of titles in this series, visit the AMS Bookstore at **www.ams.org/bookstore/**.

Photo courtesy of Marisa Gioioso

Many problems in number theory have simple statements, but their solutions require a deep understanding of algebra, algebraic geometry, complex analysis, group representations, or a combination of all four. The original simply stated problem can be obscured in the depth of the theory developed to understand it. This book is an introduction to some of these problems, and an overview of the theories used nowadays to attack them, presented so that the number theory is always at the forefront of the discussion.

Lozano-Robledo gives an introductory survey of elliptic curves, modular forms, and L-functions. His main goal is to provide the reader with the big picture of the surprising connections among these three families of mathematical objects and their meaning for number theory. As a case in point, Lozano-Robledo explains the modularity theorem and its famous consequence, Fermat's Last Theorem. He also discusses the Birch and Swinnerton-Dyer Conjecture and other modern conjectures. The book begins with some motivating problems and includes numerous concrete examples throughout the text, often involving actual numbers, such as 3, 4, 5, $\frac{3344161}{747348}$, and $\frac{22440351770433696992455751309066748631609484720 41}{89123322689288595880255351789671635700164808 30}$.

The theories of elliptic curves, modular forms, and L-functions are too vast to be covered in a single volume, and their proofs are outside the scope of the undergraduate curriculum. However, the primary objects of study, the statements of the main theorems, and their corollaries are within the grasp of advanced undergraduates. This book concentrates on motivating the definitions, explaining the statements of the theorems and conjectures, making connections, and providing lots of examples, rather than dwelling on the hard proofs. The book succeeds if, after reading the text, students feel compelled to study elliptic curves and modular forms in all their glory.

ISBN 978-0-8218-5242-2

9 780821 852422

STML/58

For additional information and updates on this book, visit

www.ams.org/bookpages/stml-58