ELLIPTIC CURVES AND THEIR APPLICATIONS TO CRYPTOGRAPHY

An Introduction

Andreas Enge

KLUWER ACADEMIC PUBLISHERS

ELLIPTIC CURVES AND THEIR APPLICATIONS TO CRYPTOGRAPHY
An Introduction

Andreas Enge

KLUWER ACADEMIC PUBLISHERS

ELLIPTIC CURVES AND THEIR APPLICATIONS TO CRYPTOGRAPHY

An Introduction

by

Andreas Enge
Universität Augsburg, Germany

KLUWER ACADEMIC PUBLISHERS
Boston / Dordrecht / London

Distributors for North, Central and South America:
Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, Massachusetts 02061 USA
Telephone (781) 871-6600
Fax (781) 871-6528
E-Mail <kluwer@wkap.com>

Distributors for all other countries:
Kluwer Academic Publishers Group
Distribution Centre
Post Office Box 322
3300 AH Dordrecht, THE NETHERLANDS
Telephone 31 78 6392 392
Fax 31 78 6546 474
E-Mail <orderdept@wkap.nl>

Electronic Services <http://www.wkap.nl>

Library of Congress Cataloging-in-Publication Data

Enge, Andreas.
Elliptic curves and their applications to cryptography : an introduction / by Andreas Enge.
p. cm.
Includes bibliographical references and index.
ISBN 0-7923-8589-6 (alk. paper)
1. Computer security. 2. Cryptography. 3. Curves, Elliptic.
I. Title.
QA76.9.A25E544 1999
005.8 ' 2--dc21 99-33654
CIP

Printed on acid-free paper.

Printed in the United States of America

This printing is a digital duplication of the original edition.

Contents

List of Tables

List of Figures

Foreword

Since the advent of public-key cryptography in 1976 by Diffie and Hellman many public-key schemes have been devised. Almost all have what are generally considered to be "hard" mathematical problems as the basis for their security. In particular, the integer factorization problem and the discrete logarithm problem are at the heart of several of the most well known techniques.

A few public-key technologies are now being widely deployed commercially to secure such activities as electronic payment over the Internet, stock trading from pagers and cell phones, and multi-applications on smart cards. Two of the more well-known methods are the RSA scheme and the DSA (digital signature algorithm). The former bases its security on integer factorization and the latter on the discrete logarithm problem in the multiplicative group of a finite field. For both of these problems there are subexponential time algorithms, which means in practice key sizes are forced to exceed 1000 bits to attain adequate security. For many constrained environments where power, storage and bandwidth are severely limited it becomes impossible to provide public-key cryptography through these methodologies.

In 1985 Neal Koblitz and Victor Miller independently proposed elliptic curve cryptography. The security of this scheme would rest on the difficulty of the discrete logarithm problem in the group formed from the points on an elliptic curve over a finite field. To date the best method for computing elliptic logarithms is fully exponential. This translates into much smaller key sizes permitting one to deploy public-key cryptography on devices where previously it was impossible. Over the past fourteen years elliptic curve cryptography has been gaining popularity and it is now being standardized around the world by agencies such as ANSI, IEEE and ISO. Recently, in January 1999, the elliptic curve version of the DSA (called the ECDSA) became an ANSI X9.62 standard for the US financial sector.

Elliptic curve cryptography relies on the elegant but deep theory of elliptic curves over finite fields. There are, to my knowledge, very few books which provide an elementary introduction to this theory and even fewer whose motivation is the application of this theory to cryptography. Andreas Enge has written a book which addresses these issues. He has developed the basic theory in a simple but thorough manner and in an easily understandable style. I have used a preliminary version of this book from which to teach a senior undergraduate course on elliptic curve cryptography. I was so pleased with the outcome that I encouraged Andreas to publish the manuscript. I firmly believe that this book is a very good starting point for anyone who wants to pursue the theory of elliptic curves over finite fields and their applications to cryptography.

S. A. VANSTONE

Preface

"2^{ter} Schluß die Mathematik ist erschrecklich schwer."
Die Mathematik (...) ist (...) nicht allein die gewisseste und zuverlässigste aller menschlichen Wissenschaften, sondern auch gewiß die leichteste.

—Lichtenberg

During the last twenty years the invention of public key cryptosystems, in conjunction with the emerging computer technology, has opened new fields of applications for number theory and algebraic geometry, which were so far considered as the "purest" branches of mathematics. Elliptic curves are among the most promising tools in modern cryptography. This has raised new interest in the topic not only within the mathematical community, but also on the part of engineers and computer scientists, who are concerned with the implementation of new cryptosystems.

My aim is to present a textbook for those who find it hard to learn about elliptic curves from the more advanced treatments, and thus to lay the foundations for studying these more complete references. To follow this book, only undergraduate algebra is needed; the reader should basically have heard of polynomial rings, field extensions and finite fields. Even this elementary approach will eventually guide us towards questions at the forefront of current research, like the problem of counting points on elliptic curves, which was satisfactorily solved only a few years ago.

While many of the fascinating applications of elliptic curves like the factorisation of integers or primality proofs deal with curves over prime fields only, curves over fields of characteristic 2 are especially attractive in the cryptographic context. This textbook treats curves in odd and even characteristic with equal attention, referring to general arguments where possible and falling back on case distinctions where necessary.

I am grateful to Reinhard Schertz, whose enthusiastic undergraduate lectures raised my interest in elliptic curves, to Dieter Jungnickel, who suggested the topic and marvelously supervised the advance of my thesis, from which this book finally emerged, and to Leonard Charlap and David Robbins, whose excellent report on elliptic curves formed the basis for my presentation. I thank Marialuisa de Resmini and Scott Vanstone for encouraging the publication. And I am especially indebted to Dirk Hachenberger, Dieter Jungnickel, Charles Lam and Berit Skjernaa for the time they spent reading the manuscript and for their valuable comments.

I hope that the reader has as much pleasure in reading this book as I had in writing it.

ANDREAS ENGE

1 PUBLIC KEY CRYPTOGRAPHY

Wenn nicht mehr Zahlen und Figuren
Sind Schlüssel aller Kreaturen
(...) Dann fliegt vor Einem geheimen Wort
Das ganze verkehrte Wesen fort.

—Novalis

Today's widespread use of electronic networks in the economic world has raised cryptography from a speciality of the military and secret services to a topic of public interest, which concerns international organisations like the UNO ([UNCITRAL, 1998a] and [UNCITRAL, 1998b]) and the EU ([Commission of the European Communities, 1998]). Unlike conventional cryptosystems, public key cryptography is applicable on a large scale base, in principle allowing secure and authorised communication between any two persons in the world. In the following chapter we give a brief introduction to the concepts of public key cryptography and present some algorithms. We hereby focus on schemes for encryption and digital signatures which can be generalised to arbitrary groups, especially to elliptic curve groups. A comprehensive treatment of cryptographic issues is given in [Stinson, 1995] and [Menezes et al., 1997].

1.1 PRIVATE VERSUS PUBLIC KEY CRYPTOGRAPHY

Cryptography is the art of sending messages over an insecure channel, taking into account the two major issues of privacy, which means that an intruder overhearing the channel is unable to derive the contents of the messages, and of authenticity, which means that the identity of the sender and the integrity of the transmitted messages can be verified. We first concentrate on how privacy can be achieved and discuss the topic of authenticity in Section 1.4.

Nowadays, the channel is usually an electronic network and the messages may be written text like an e-mail or a contract, but also any other data such as a technical blueprint, software or the human voice in digitalised form. To secure the content of a message against an intruder into the network, the raw data is transformed into a random looking form by using mathematical operations.

So, in a first step the message has to be coded into a mathematical object. Usually it is broken into blocks of fixed length, and each block is transformed into an integer or a bit sequence. This coding step is merely of technical importance and does not provide any privacy yet. We are not concerned with details of this process and simply refer to the encoded blocks as "messages" again.

In a second step each block is transformed in such a way that an unauthorised person is unable to retrieve the original block. It is sent to the intended recipient, who inverts the transformation and recomposes the original message. Roughly speaking the algorithms for transforming blocks form a *cryptosystem.*

More formally, a cryptosystem consists of three finite sets $\mathcal{M}$, $\mathcal{C}$ and $\mathcal{K}$, the *message*, *ciphertext* and *key space*, and of a family of *encryption functions* $f_k : \mathcal{M} \to \mathcal{C}$, indexed with the elements k of $\mathcal{K}$. In the terminology of the previous paragraph, $\mathcal{M}$ is the set of message blocks and $\mathcal{C}$ the set of transformed blocks. To make encryption possible, f_k must be effectively computable, and to allow decryption, f_k must be injective. Very often $\mathcal{M} = \mathcal{C}$ and the f_k are bijections.

When Kevin wishes to send a secret message $m \in \mathcal{M}$ to Laura he chooses a *key* $k \in \mathcal{K}$. He then computes the *ciphertext* $c = f_k(m)$ and sends it to Laura over a possibly insecure channel. Laura must be able to apply the inverse *decryption function* f_k^{-1} to c to obtain the original message $m = f_k^{-1}(c)$.

In a conventional, so-called *private key* cryptosystem, the knowledge of f_k or f_k^{-1} is equivalent to that of k. This means that a person who has the ability of encryption can decrypt ciphertexts with the same effort and vice versa. Thus two major drawbacks of these methods can be stated:

Key distribution problem: Any two persons who wish to communicate in private have to agree on a common key. Hence they need a secure channel to exchange the key beforehand, i.e. they must meet personally, use a trusted courier or any other secure method. This secure channel is usually much more expensive than the insecure one used to send the subsequent messages. To keep a high level of security the key should be changed regularly, which increases the operating cost of the system. Moreover, in a network of many persons the overall number of keys grows quadratically with the number of participants, resulting in a possibly unmanageable amount of keys.

Signature problem: To make agreements over electronic networks legally binding the recipient of a secret message must be able to prove the identity of the sender to a third party, e.g. a judge. Since in a conventional cryptosystem someone who can decrypt a ciphertext can also encrypt arbitrary messages, it is no problem for a recipient to forge messages of his choice.

In 1976, Diffie and Hellman proposed a solution to these problems which has revolutionised cryptography ([Diffie and Hellman, 1976]). Their approach is based on *one-way functions*, or, to be more precise, *trap-door one-way functions*. Imagine that for each key k an encryption function f_k is chosen such that, even knowing f_k, it is computationally infeasible to determine f_k^{-1}. Then, Kevin can publish his encryption function f_k, the so-called *public key*, and anyone who wishes to send him a secret message can do so, including Laura. Of course this system is of little use, because even Kevin cannot decipher the messages he has received (but note that such systems are widely used for authentication purposes where only the encrypted version of a password has to be stored). This is why the trap-door is needed: For a trap-door one-way function f_k it is easy to determine f_k^{-1}, provided that the *secret key* or *trap-door* k is known. With the extra knowledge of k, Kevin can now decrypt his messages.

This procedure solves both problems described above: To distribute the keys it is no more necessary to dispose of a secure channel. On the contrary, the *public* keys should be published in a prominent place to allow secret communication between strangers, e.g. a mail order house and its clients. To sign a message m which Kevin wishes to send to Laura, maybe his part of a contract, he can first *decrypt* the message with his private function f_k^{-1} and send the pair $(m, f_k^{-1}(m))$ to Laura. Since Kevin is the only person who knows the algorithm f_k^{-1}, Laura can prove to any third party that the pair she has received comes from Kevin by comparing m with $f_k(f_k^{-1}(m))$. If additional privacy is desired the method can be applied to $c = f_l(m)$ instead of m, with f_l being Laura's public key.

Beside the rather informal way of describing trap-door one-way functions using terms like "easy to compute" or "computationally infeasible" ([Diffie and Hellman, 1976], p. 648) there does not seem to exist a commonly accepted definition. The reason becomes clear when we try to formulate some minimal requirements for such a function. As is common in complexity theory we assume that "easy to compute" means "computable by a deterministic algorithm in polynomial time". Then it should be possible to determine f_k and f_k^{-1} from k in polynomial time, and any $f_k(m)$ and $f_k^{-1}(c)$ should be computable in polynomial time, if the algorithms f_k and f_k^{-1} are known. On the other hand, determining f_k^{-1} or any single $f_k^{-1}(c)$ from the knowledge of f_k (and thus of the values $f_k(m)$ for *any chosen* m) should be impossible in polynomial time. Even with these reasonable demands it is not sure whether a trap-door one-way function exists: There is an obvious non-deterministic polynomial method for obtaining $f_k^{-1}(c)$, namely guessing the message m and testing for $f_k(m) = c$. Now, if the sets of problems which are solvable in polynomial time by a deterministic or non-deterministic algorithm coincide, i.e. , if $P = NP$, then there is no trap-door one-way function. The question whether $P = NP$, however, is one of the central open problems in complexity theory.

In practice a trap-door one-way function is a function which is computed in "reasonable" time, even with little computing power, while its inverse cannot be computed in reasonable time, even with the best algorithms known and with high computing power. Of course such a definition is subject to the security needs of the users, and a secret service will apply different standards than a person sending holiday postcards. Several trap-door one-way functions have been proposed in literature, and we present some (but by far not all) of them in the following sections.

The application of the encryption function usually consists of elementary operations in the underlying mathematical objects $\mathcal{M}$ or $\mathcal{C}$, and since it has to be inverted by the intended recipient of the message, it is quite natural to require the same property for the elementary operations. Consequently the standard setting is to choose $\mathcal{M} = \mathcal{C}$ as a group, or as a monoid in which only very few elements are not invertible (usually the failure of a trial inversion leads to cracking the system and thus should occur with a merely negligible probability). We restrict our presentation to the group case and formulate the algorithms of the following sections in general (multiplicative) group notation, so that they specialise naturally to elliptic curve groups.

1.2 DIFFIE–HELLMAN KEY EXCHANGE

The scheme presented by Diffie and Hellman is not yet a full public key cryptosystem, they classify it as "public key distribution scheme" ([Diffie and Hellman, 1976], p. 468). The idea is to exchange pieces of information over an insecure channel so that both parties share a common secret afterwards, whereas an intruder who happens to catch the partial information cannot reconstruct the secret. This secret can then be used as key in a conventional cryptosystem. The details are as follows:

1. Kevin and Laura publicly choose a cyclic group G and a generator α of G. (In the original paper G is the multiplicative group of a finite field.)

2. As private keys Kevin and Laura choose random integers k and l, respectively. They compute α^k and α^l, respectively, and exchange these values.

3. Now both of Kevin and Laura can compute α^{kl} via

   $$\alpha^{kl} = (\alpha^k)^l = (\alpha^l)^k$$

   using the information they received from one another and their private keys. α^{kl} is the shared secret.

Notice that α can be raised very efficiently even to high powers by the square and multiply algorithm.

Algorithm 1.1 (Square and multiply) *Let α be an element of a group and k a natural integer. The following algorithm computes $\gamma = \alpha^k$ with $O(\log k)$ group operations.*

1. *Let $\gamma = 1$.*

2. *Repeat the following step until $k = 0$.*

3. *If k is odd, replace k by $k - 1$ and γ by $\gamma\alpha$. Now k is even in any case. Replace k by $\frac{k}{2}$ and α by α^2.*

Proof: During the execution of the algorithm the value of $\gamma\alpha^k$ is invariant, whence γ contains the desired power as soon as $k = 0$. This proves the correctness of the method. If the binary representation of k has r bits, i.e. $2^{r-1} \leq k < 2^r$, and $s \leq r$ of these bits are 1, then the algorithm requires exactly $r-1$ squarings and $s-1$ further multiplications. This proves the assertion on the complexity. □

When the group is abelian, changing the algorithm to additive notation yields the double and add algorithm, which was used by the Ancient Egyptians to multiply integers, see [Gillings, 1972].

An eavesdropper who overhears the communication in a Diffie–Hellman key exchange and wants to reconstruct the secret has the task of computing α^{kl} from α, α^k and α^l. This problem is known as the *Diffie–Hellman problem*. One obvious approach is to compute k from α^k, which is the *discrete logarithm* of α^k to the base α. Note that k is determined only up to the order of G, but that the knowledge of any k' with $\alpha^{k'} = \alpha^k$ suffices to compute $\alpha^{kl} = (\alpha^l)^{k'}$.

It is still unknown whether the Diffie–Hellman problem and the discrete logarithm problem are computationally equivalent, but this is widely believed to be true. (In fact, for a large class of finite groups, Maurer and Wolf proved the equivalence, see [Maurer and Wolf, 1996].) If so, then the security of the Diffie–Hellman system would rely on the difficulty of the discrete logarithm problem, which in turn depends on the representation of the group: If $G = (\mathbb{Z}_n, +)$ is the cyclic group of integers modulo n represented by $\{0, \ldots, n-1\}$ and $\alpha = 1$, then the problem becomes trivial. We will come back to discrete logarithms in Chapter 4 and note for the time being that no polynomial algorithm is known for the multiplicative group of a finite field, given in standard polynomial or normal basis representation.

1.3 ELGAMAL CRYPTOSYSTEM

Based on the Diffie–Hellman key exchange scheme, ElGamal proposed a full public key cryptosystem in [ElGamal, 1985]. Again a cyclic group G with generator α is fixed; $\mathcal{M} = G$ and $\mathcal{C} = G \times G$. Each participant chooses a private key $a \in \mathbb{Z}$ and publishes α^a. Suppose that Kevin wishes to send a message m to Laura.

1. Kevin chooses a random integer k and looks up Laura's public key α^l.

2. He computes $\alpha^{kl} = (\alpha^l)^k$ and sends the pair $(\alpha^k, m\alpha^{kl})$ to Laura.

3. Laura computes $\alpha^{kl} = (\alpha^k)^l$ with her private key and retrieves m.

The scheme is obviously equivalent to the Diffie–Hellman key exchange; it just avoids one transaction by storing α^l in a public place. A drawback is the *message expansion* by a factor of 2: To communicate the information contained in one group element, two group elements have to be transmitted. This can be avoided if Kevin uses his public key α^k for all transactions so that it suffices to send α^k once and for all. This simplification, however, opens a security gap: If an eavesdropper happens to know one message–ciphertext pair $(m_1, m_1\alpha^{kl})$, he is able to retrieve a second message m_2 from its ciphertext $m_2\alpha^{kl}$ by computing

$$m_2 = m_1 \frac{m_2\alpha^{kl}}{m_1\alpha^{kl}}.$$

A reasonable compromise would be to choose a different key k for each "session", e.g. an e-mail transfer, which usually consists of several consecutive messages $m_1, m_2, \ldots, m_n$. The encrypted e-mail would then be composed of α^k and of $m_1\alpha^{kl}$, $m_2\alpha^{kl}$, $\ldots, m_n\alpha^{kl}$, which reduces the message expansion factor to $1 + \frac{1}{n}$.

Notice that in contrast to the general concept of Section 1.1 there is a slight asymmetry in the scheme which makes signatures impossible: Laura is able to produce valid ciphertexts for any message m since she can either choose an arbitrary value for α^k or knows α^k as Kevin's public key. ElGamal suggested a different signature scheme which is presented in the next section.

1.4 SIGNATURE SCHEMES

We present some important examples of signature schemes which fit the context of the previous sections. Suppose that Kevin wishes to send a signed message m to Laura. Then he adds a "signature" to the message which depends on his secret key k and on the message itself. Checking the signature, Laura can prove that Kevin is the real sender and that the message has not been altered during the transmission.

ElGamal Signature Scheme

Together with his encryption technique, ElGamal proposed the following signature scheme in [ElGamal, 1985], which is again stated for an arbitrary finite cyclic group G with generator α. It is based on the Diffie–Hellman problem in G. Suppose that $g : \mathcal{M} = G \to \{0, \ldots |G| - 1\}$ is an efficiently computable bijection, and that $m \in G$ is the message to be signed. Kevin's private and public keys are k and α^k, respectively.

Signing a message

1. Kevin chooses a random integer k', coprime to $|G|$, and computes $r = \alpha^{k'}$.

2. He solves the congruence

$$g(m) \equiv kg(r) + k's \pmod{|G|}. \tag{1.1}$$

Since k' and $|G|$ are coprime, there is a unique solution $s \in \{0, \dots |G| - 1\}$.

3. The signature is the pair $(r, s) \in G \times \mathbb{Z}_{|G|}$, and it is sent together with m to Laura.

Verifying a signature

1. Laura computes $\alpha^{g(m)}$ and $\alpha^{kg(r)+k's} = (\alpha^k)^{g(r)} r^s$ from Kevin's public key α^k and from m, r and s.

2. If the two values agree, she can assume the validity of the signature.

For an analysis of the security of the system, see the original paper, pp. 470–471. It is a crucial point to choose distinct k' for different messages, since otherwise only two messages and their respective signatures may suffice to solve the system of linear equations formed by (1.1) for k' and Kevin's private key k.

Again a major disadvantage of the scheme is the expansion of transferred data — a signed message is thrice as long as the message itself. This overload can be reduced by means of a *hash function*. In practice, data communicated during a session consists of several blocks $m_1, m_2, \dots, m_n$. A hash function is a function

$$h : \mathcal{M}^{(\mathbb{N})} \to \mathcal{H},$$

where $\mathcal{M}^{(\mathbb{N})}$ denotes the set of finite sequences with entries in $\mathcal{M}$ and $\mathcal{H}$ is a finite set. Instead of signing each block m_i separately one can sign only the hash value $h(m_1, m_2, \dots, m_n)$. In order to prevent forgery of signatures, h should be a one-way function.

Digital Signature Standard

In 1994, the National Institute for Standards and Technology (NIST) of the USA announced a standard for digital signatures (DSS) which is obligatory for US governmental agencies (see [NIST, 1994]). Besides prescribing the use of a specific hash function, the DSS fixed an algorithm for digital signatures (DSA). (The updated current version also allows RSA signatures, see [NIST, 1998].) The general setting is as follows:

1. p is a prime with $2^{L-1} < p < 2^L$ where

$$L \in \{512, 576, 640, 704, 768, 832, 896, 960, 1024\}.$$

2. q is a prime divisor of $p - 1$ with $2^{159} < q < 2^{160}$.

3. $\alpha \in \mathbb{F}_p$ is a generator of the unique subgroup of $\mathbb{F}_p^\times$ of order q. The signature algorithm is based on the discrete logarithm problem in $\langle\alpha\rangle$.

4. The function

$$g : \mathbb{F}_p^\times = \{1, \ldots, p-1\} \to \{0, \ldots, q-1\}$$

is the reduction modulo q:

$$g(\alpha) \equiv \alpha \pmod{q} \quad \text{for } \alpha \in \{1, \ldots, p-1\}.$$

5. $h : \mathcal{M}^{(\mathbb{N})} \to \mathbb{Z}$ is the hash function determined by the Secure Hash Standard (SHS) as specified in [NIST, 1995].

Kevin's private key is an integer k with $0 < k < q$, his public key is α^k. Suppose that m is the message to be signed.

Signing a message

1. Kevin chooses a random integer k' with $0 < k' < q$ and computes $r = \alpha^{k'}$ and $g(r)$.

2. He solves the congruence

$$h(m) \equiv -kg(r) + k's \pmod{q} \tag{1.2}$$

with $0 < s < q$.

3. The signature is the pair $(g(r), s) \in \mathbb{Z}_q \times \mathbb{Z}_q$.

Verifying a signature

1. Laura solves $ws \equiv 1 \pmod{q}$ with $0 < w < q$ and computes the values $u_1 \equiv h(m)w \pmod{q}$, $u_2 \equiv g(r)w \pmod{q}$ with $0 \leq u_1, u_2 < q$ and $v = g(\alpha^{u_1}(\alpha^k)^{u_2})$.

2. If $v = g(r)$, Laura can assume the validity of the signature.

   Let us verify that Laura accepts correct signatures. Note that (1.2) implies

$$\alpha^{u_1}(\alpha^k)^{u_2} = \alpha^{h(m)w+kg(r)w} = \alpha^{k'sw} = \alpha^{k'}.$$

Thus

$$v = g(\alpha^{u_1}(\alpha^k)^{u_2}) = g(r).$$

The reduction modulo q of elements of $\mathbb{F}_p^\times$, which is in fact an arbitrary mapping $g : \mathbb{F}_p^\times \to \{0, \ldots, q-1\}$ because it is incompatible with the multiplicative structure of $\mathbb{F}_p$, is intended to reduce the signature length from $2L$ bits to 320 bits.

The generalisation of the signature scheme to arbitrary finite cyclic groups $G = \langle\alpha\rangle$ is straightforward and reveals that the DSA is equivalent to ElGamal's algorithm in a special context: The signing procedure is exactly the same, whereas the verification process is slightly modified and requires only the knowledge of $g(r)$ and not that of r.

1.5 STANDARDS

Due to the increasing popularity of public key cryptography, different national and international organisations are working on standardising the algorithms used and their parameters. We briefly mention some publicly accessible documents.

The Institute of Electrical and Electronics Engineering is preparing a comprehensive standard for public key cryptography, including different cryptographic primitives, signature schemes and key agreement protocols ([IEEE, 1998]). The American National Standards Institute is issuing a series of standards intended for financial services. [ANSI, 1998a] specifies key agreement protocols based on the discrete logarithm problem in finite fields, like the Diffie–Hellman key exchange. Of special interest in the context of this book are [ANSI, 1999] and [ANSI, 1998b]. The first standard covers key exchange schemes bases on elliptic curves, while the second one specifies ECDSA, the elliptic curve analogue of DSA.

We already mentioned in Section 1.4 that the US government actually prescribes the use of DSA or RSA for electronic signatures. The European Union will be more flexible. It focuses on the legal implications of electronic signatures, and its regulations are intended to assure the functioning of certification services, which are needed to undoubtedly link a person and corresponding signatures. But it does not plan to fix a specific technology: "While there is much discussion and work on digital signature technologies which employ public-key cryptography, a Directive at the European level should be technology-neutral and should not focus only on these kinds of signatures. Since a variety of authentication mechanisms is expected to develop, the scope of this Directive should be broad enough to cover a spectrum of 'electronic signatures', which would include digital signatures based on public-key cryptography as well as other means of authenticating data." ([Commission of the European Communities, 1998], p. 3). However, the European Union expects that international standards will emerge and encourages the industry to adopt these standards. For a condensed overview of legislative initiatives taken in different European countries, see also [Commission of the European Communities, 1998].

In their attempt to define an international legal framework for electronic signatures, the United Nations follow a similar strategy of not focusing on specific technological issues, see [UNCITRAL, 1998a] and [UNCITRAL, 1998b].

2 THE GROUP LAW ON ELLIPTIC CURVES

Das Erst' wär' so, das Zweite so,
Und drum das Dritt' und Vierte so;
Und wenn das Erst' und Zweit' nicht wär',
Das Dritt' und Viert' wär' nimmermehr.

—Goethe

Elliptic curves can be equipped with an efficiently computable group law, so that they are suited for implementing the cryptographic schemes of the previous chapter, as suggested first in [Koblitz, 1987] and [Miller, 1986]. They are particularly appealing because they achieve the same level of security as a finite field based cryptosystem with much shorter key lengths, which results in a faster encryption and decryption process. Our aim in this chapter is to prove the group law.

After presenting the necessary definitions we show that there is an intuitive geometric composition law on an elliptic curve, involving lines and their intersection points with the curve. Some elementary computations result in simple algebraic formulae which are suited for computer implementations. The composition law fulfils all group axioms, but strange enough, its associativity is hard to prove. It can be shown in various ways:

The obvious approach is brute force computation, the explicit algebraic formulae for adding two points on a curve being given. Unfortunately there are several formulae, depending on the position of the points to be added, and so an awful lot of case distinctions is needed. What is worse, the proof does not reveal anything about the underlying algebraic and geometric structures and is not only extremely tedious, but also extremely uninstructive. This seems to have deterred most authors, for, to my knowledge, this approach cannot be found in any publication.

Some authors concentrate on elliptic curves over the complex numbers, where the additional analytic structure accounts for particular properties, see [Koblitz, 1993], [Lang, 1978] or [Lang, 1987]. But for implementational reasons we are mainly interested in curves over finite fields, to which the analytic proofs do not apply. Hence in this book we concentrate on purely algebraic approaches, which work over any field. It is instructive, however, to relate our algebraic findings to their analytic counterparts, and the reader is invited to take a closer look at the books mentioned above.

Fulton presents a beautiful geometric proof in his book on algebraic curves (see [Fulton, 1969], p. 125) after developing some general theory. The same proof is reported in [Husemöller, 1987], Chapter 3. Other arguments use the Riemann–Roch theorem, which is presented at the end of Fulton's book. These approaches are ideal for specialists in algebraic geometry, in which case the standard references are [Silverman, 1986] and [Silverman, 1994]. However, elliptic curves are still quite "simple" from the algebraic-geometric point of view and can be understood without knowing much of abstract algebraic geometry.

In this chapter we follow Charlap's and Robbin's elementary proof (see [Charlap and Robbins, 1988]). On one hand, we explain the basic notions of the theory of algebraic curves, so that the reader gets an introduction to this topic. On the other hand, it is our aim to keep this exposition as elementary and concrete as possible. So we specialise all results to the case of elliptic curves, where many of them can be proved by explicit computations or more elementary arguments than in the general case. Unlike Charlap and Robbins we consistently use the projective point of view when working with the infinite point $\mathcal{O}$, which appears naturally in this setting, and thus avoid seemingly artificial constructions. Furthermore we present a generalised version of the proof, which covers fields of any characteristic, including the case of characteristic 2, which is highly relevant for cryptography.

THROUGHOUT THIS CHAPTER, LET K BE AN ALGEBRAICALLY CLOSED FIELD.

2.1 AFFINE PLANE CURVES

In this section we introduce the basic definitions and notations related to affine curves. Our main concern during the following sections will be to fill these concepts with life and to specialise them to the case of elliptic curves.

Definition 2.1 *The set $K \times K$ is called the* affine plane over K *and denoted by $A^2(K)$.*

Definition 2.2 *An* affine plane curve over K *is the set of zeros of an irreducible polynomial* $C \in K[X,Y]$ *in the affine plane, i.e.* $\{(x,y) \in A^2(K) : C(x,y) = 0\}$.

Example. For $K = \mathbb{R}$, the curves $D = Y^2 - (X^3 + X^2)$ and $E = Y^2 - (X^3 - X)$ are drawn in Figure 2.1.

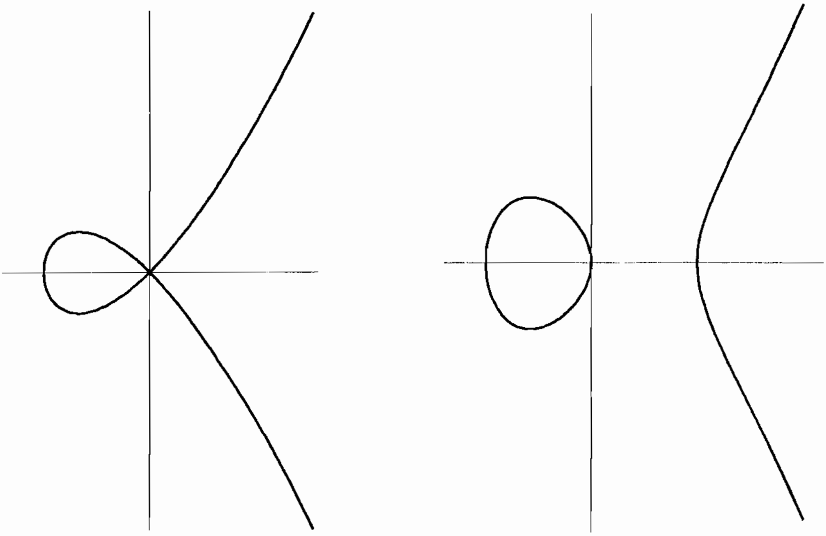

Figure 2.1. The curves $Y^2 = X^3 + X^2$ and $Y^2 = X^3 - X$

Since K is algebraically closed, any curve contains an infinite number of points: For all $x \in K$, the equation $C(x,Y)$ has at least one zero y in K resulting in a point $P = (x,y)$ of the curve. For notational convenience we denote a curve by the same letter as its defining polynomial and call it "the curve defined by C", "the curve with equation C" or simply "the curve C".

Definition 2.3 *Let C be a curve and $P = (x,y)$ a point on C. Then P is* singular *on C if* $\frac{\partial C}{\partial X}(x,y) = \frac{\partial C}{\partial Y}(x,y) = 0$. *A* singular curve *is a curve with at least one singular point.*

Example. Since $\frac{\partial D}{\partial X}(0,0) = -(3X^2 + 2X)|_{X=0} = 0$ and $\frac{\partial D}{\partial Y}(0,0) = 2Y|_{X=0} = 0$, the origin is singular on D. On E, however, the origin is non-singular because $\frac{\partial E}{\partial X}(0,0) = (-3X^2 + 1)|_{X=0} = 1 \neq 0$. Geometrically non-singularity implies that there is a unique tangent line. E. g., the unique tangent at E in the origin is the vertical line. As can be seen from the figure, D has two distinct tangents in its singular point $(0,0)$, namely $Y = X$ and $Y = -X$. The origin is called a *node* in this case. Another possibility would be a *cusp*, i.e. a point with only one, but multiple tangent. We will come back to tangents (and give a rigorous definition) in Section 2.9 about lines.

We are interested in polynomial functions on a curve. Obviously two polynomials f and g agree as functions on a curve C if they differ only by a multiple of C, i.e. $C \mid f - g$. In fact the converse is also true, an assertion we will establish later for the special case of elliptic curves (see Section 2.2). This leads to the following definition:

Definition 2.4 *The* coordinate ring *of a curve C is $K[C] := K[X,Y]/(C)$.*

To simplify the notation we denote the residue classes of X and Y in $K[C]$ again by X and Y; the meaning will always be clear from the context. As C is irreducible, $K[C]$ is an integral domain.

Definition 2.5 *The field of fractions of $K[C]$ is called the* field of rational functions on C *and is denoted by $K(C)$.*

While it is always possible to evaluate a polynomial of the coordinate ring at a given point of the curve by substituting the X- and Y-coordinates, this need not be the case for a rational function because zero denominators may occur. Note that although $K[X,Y]$ is a unique factorisation domain this is usually not the case for $K[C]$.

Definition 2.6 *Let P be a point on C. A rational function $r \in K(C)$ is called* regular *or* defined *at P if there exist $f, g \in K[C]$ such that $r = \frac{f}{g}$ and $g(P) \neq 0$. The value of r at P is then $r(P) = \frac{f(P)}{g(P)}$. The ring of all rational functions regular at P is called the* local ring of C at P *and is denoted by $\mathcal{O}_P(C)$. We sometimes write $r(P) = \infty$ if r is not regular at P.*

Notice that the value of a regular rational function at a point is *well defined*, i.e. independent of the representation as a quotient of two polynomials. Assume that $r = \frac{f_1}{g_1} = \frac{f_2}{g_2}$ with f_1, g_1, f_2, $g_2 \in K[C]$ and $g_1(P)$, $g_2(P) \neq 0$. Then $f_1 g_2 = f_2 g_1$ in $K[C]$, which means that there is a polynomial $h \in K[X,Y]$ such that $f_1 g_2 - f_2 g_1 = hC$ in $K[X,Y]$. Hence, $f_1(P)g_2(P) - f_2(P)g_1(P) = h(P)C(P) = h(P) \cdot 0 = 0$, which implies $\frac{f_1(P)}{g_1(P)} = \frac{f_2(P)}{g_2(P)}$.

It is then easy to verify that $\mathcal{O}_P(C)$ is a ring, and indeed it is a local ring: Its *units*, i.e. its invertible elements, are $\mathcal{O}_P(C)^\times = \left\{\frac{f}{g} : f(P),\, g(P) \neq 0\right\}$ and its unique maximal ideal is $\mathfrak{m}_P(C) = \left\{\frac{f}{g} : g(P) \neq 0,\, f(P) = 0\right\}$.

Example. Let $E : Y^2 = X^3 - X$ be the right hand side curve of Figure 2.1, and consider the rational function $r = \frac{X}{Y}$. In $P = (0,0)$ we have $X(P) = Y(P) = 0$; both the numerator and the denominator of r are zero in P. However, it is possible to choose a different representation for r:

$$r = \frac{X}{Y} = \frac{XY}{Y^2} = \frac{XY}{X^3 - X} = \frac{Y}{X^2 - 1}.$$

Now $Y(P) = 0$ and $(X^2 - 1)(P) = -1 \neq 0$, which means that r is regular at P with $r(P) = \frac{0}{-1} = 0$.

Let $s = \frac{1}{r} = \frac{Y}{X}$. Then s is not regular at P. Otherwise $s(P)$ would be an element of K, and we could compute

$$1 = 1(P) = (rs)(P) = r(P)s(P) = 0\,s(P) = 0,$$

a contradiction.

2.2 AFFINE ELLIPTIC CURVES

In this section we introduce our main objects of study, the elliptic curves. We have a closer look at the occurring coordinate rings and fields of rational functions, which are of a special form in this case.

Definition 2.7 *An* affine Weierstraß equation *over K is an equation of the form*

$$E : Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6$$

with a_1, a_3, a_2, a_4, $a_6 \in K$. The following quantities are related to E:

$$\begin{aligned}
b_2 &= a_1^2 + 4a_2 \\
b_4 &= 2a_4 + a_1a_3 \\
b_6 &= a_3^2 + 4a_6 \\
b_8 &= a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2 \\
c_4 &= b_2^2 - 24b_4 \\
\Delta &= -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6 \\
j &= \frac{c_4^3}{\Delta} \quad \text{for } \Delta \neq 0
\end{aligned}$$

Δ is called the discriminant *of E, j its* j-invariant. *A non-singular Weierstraß equation defines an* elliptic curve.

The quantities b_2, b_4, b_6, b_8 and c_4 are only needed to shorten the definition of Δ and j. In Section 2.4 we will see that Δ determines whether the Weierstraß equation is singular or not. We will not have to deal with j too often during this exposition; the normal forms of Section 2.3 depend on it, and Section 3.11 gives another hint at its relevance. It is notationally convenient not to distinguish between an elliptic curve, its defining equation or its defining polynomial, so henceforth we denote these three objects by the common letter E. When speaking of E as a polynomial, we mean $E = Y^2 + a_1XY + a_3 - (X^3 + a_2X^2 + a_4X + a_6)$.

Let us verify that the definition of an elliptic curve does not collide with the general definition of a curve of Section 2.1, i.e. that any Weierstraß equation is irreducible.

Recall the degree function on $K(X)$, which is given by

$$\deg \frac{f}{g} = \deg f - \deg g \quad \text{for } f, g \in K[X].$$

Then for $r, s \in K(X)$ we have the usual rules

$$\begin{aligned} \deg(rs) &= \deg r + \deg s, \\ \deg \frac{1}{r} &= -\deg r \text{ and} \\ \deg(r+s) &\leq \max\{\deg r, \deg s\} \quad \text{with equality if } \deg r \neq \deg s. \end{aligned}$$

As E is monic in Y and $K[X]$ a unique factorisation domain, E is irreducible in $K[X, Y]$ if and only if it is irreducible in $K(X)[Y]$. So if E were reducible it could be written as $(Y + r)(Y + s)$ where $r, s \in K(X)$. Comparing the coefficients of Y it follows that $r + s = a_1X + a_3$ and $rs = -(X^3 + a_2X^2 + a_4X + a_6)$, so $\deg(r + s) \leq 1$, and $\deg(rs) = \deg r + \deg s = 3$ is odd. Hence $\deg r \neq \deg s$, and

$$1 \geq \deg(r+s) = \max\{\deg r, \deg s\} \geq \frac{1}{2}(\deg r + \deg s) = \frac{3}{2},$$

a contradiction.

So E is irreducible, and $K[E] = K[X, Y]/(E)$ and $K(E)$, the field of fractions of $K[E]$, can be defined as in the previous section. Every element of $K[E]$ gives rise to a polynomial function on the points of E. However, it is a priori unclear if two distinct elements of $K[E]$ differ also as functions on the curve. Our aim is to prove that this is indeed the case. It is sufficient to show that if $f \in K[E]$ is zero as a function on E, i.e. $f(P) = 0$ for all $P \in E$, then it is the zero element of $K[E]$. This is true for any curve, but the proof in the case of elliptic curves is much less involved.

To this purpose another view of $K(E)$ is helpful. Clearly $K[X]$ is a subring of $K[E]$, so $K(X)$ is a subfield of $K(E)$. Adjoining a variable Y to $K(X)$ and reducing modulo E one obtains $L := K(X)[Y]/(E)$. As E is irreducible over $K(X)$, L is a field. It is clear that $K[E] \subseteq L \subseteq K(E)$, so we conclude that $L = K(E)$. Thus $K(E)$ is a quadratic extension of $K(X)$, and the unique non-trivial automorphism of $K(E)$ fixing $K(X)$ elementwise is the conjugation assigning to Y the second root of E over $K(X)$, namely $\overline{Y} = -Y - a_1X - a_3$. The conjugate of f, where each occurrence of Y is replaced by $\overline{Y}$, is denoted by $\bar{f}$. A corresponding conjugation can be defined on the points $P = (x, y) \in E$ by letting $\overline{P} = (\overline{X}, \overline{Y})(P) = (x, -y - a_1x - a_3)$, so that $\bar{f}(P) = f(\overline{P})$.

Definition 2.8 *The* norm *and* trace *functions on $K(E)$ are defined by the usual norm and trace of the field extension $K(E)/K(X)$:*

$$\begin{aligned} \mathrm{N} &: K(E) \to K(X), \quad f \mapsto f\bar{f} \\ \mathrm{Tr} &: K(E) \to K(X), \quad f \mapsto f + \bar{f} \end{aligned}$$

Example. For the *coordinate functions* X and Y we have $\mathrm{N}(X) = X^2$, $\mathrm{Tr}(X) = 2X$, $\mathrm{N}(Y) = -(X^3 + a_2X^2 + a_4X + a_6)$ and $\mathrm{Tr}(Y) = -(a_1X + a_3)$.

More generally, any $f \in K(E)$ has a representation $f = v + Yw$ with v, $w \in K(X)$. Then $\overline{f} = v + \overline{Y}w$, and $\mathrm{N}(f) = v^2 + \mathrm{Tr}(Y)vw + \mathrm{N}(Y)w^2$ and $\mathrm{Tr}(f) = 2v + \mathrm{Tr}(Y)w$.

The norm will be a useful tool throughout this chapter because it reduces polynomials in two variables to polynomials in only one variable, which are usually easier to study. Especially it can be used for proving that $K[E]$ is the ring of polynomial functions on E: Let $f \in K(E)$ be the zero function on E. Then $\mathrm{N}(f)$ is also zero on E. But $\mathrm{N}(f) \in K(X)$, and any $x \in K$ occurs as the X-coordinate of a point on E, so $\mathrm{N}(f) = 0$ and obviously $f = 0$.

2.3 VARIABLE CHANGES AND NORMAL FORMS

While our more explicit approach to elliptic curves has the advantage of needing a smaller burden of general theory it has the drawback of requiring heavier computation. The computational effort can be reduced considerably by transforming the equation of an elliptic curve such that some of the coefficients a_i become zero. Of course we admit only transformations which do not change the "behaviour" of the curve.

Definition 2.9 *Two curves defined by Weierstraß equations*

$$\begin{aligned} E &: Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6, \\ E' &: Y^2 + a_1'XY + a_3'Y = X^3 + a_2'X^2 + a_4'X + a_6' \end{aligned}$$

are isomorphic *if E' can be obtained from E by a change of variables of the form*

$$\psi : \begin{pmatrix} X \\ Y \end{pmatrix} \mapsto \begin{pmatrix} u^2X + r \\ u^3Y + u^2sX + t \end{pmatrix} = \begin{pmatrix} u^2 & 0 \\ u^2s & u^3 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} r \\ t \end{pmatrix}$$

where $u \in K^\times$, r, s, $t \in K$ (and dividing the resulting equation by u^6). The corresponding transformation ψ is referred to as an admissible change of variables.

Example. The conjugation of the coordinate functions

$$(X, Y) \mapsto (\overline{X}, \overline{Y}) = (X, -Y - a_1X - a_3)$$

is an admissible change of variables with $u = -1$, $r = 0$, $s = -a_1$ and $t = -a_3$. Moreover it is an involution, i.e. its own inverse.

In fact the above definition should be a theorem: The admissible changes of variables represent the only isomorphisms in the category of algebraic subvarieties of the affine plane which preserve the form of the Weierstraß equation. As we are not going to study algebraic geometry in its details, some plausible arguments for this definition must suffice.

First let us verify that the definition makes sense, i.e. that being isomorphic is an equivalence relation on Weierstraß equations. Clearly, letting $u = 1$ and

$r = s = t = 0$, the identity $\begin{pmatrix} X \\ Y \end{pmatrix} \mapsto \begin{pmatrix} X \\ Y \end{pmatrix}$ is an admissible change of variables, so the relation is reflexive. The inverse of ψ is given by

$$\psi^{-1} : \begin{pmatrix} X \\ Y \end{pmatrix} \mapsto \begin{pmatrix} u^{-2} & 0 \\ -u^{-3}s & u^{-3} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} - \begin{pmatrix} u^{-2}r \\ u^{-3}(t - rs) \end{pmatrix},$$

which is clearly of the same form, so the isomorphism relation is symmetric. Some simple computations, which are left as an exercise to the reader, show that the relation is also transitive.

On the other hand, curves are defined as point sets, and the term "isomorphic" is justified because to the admissible change of variables ψ correspond inverse bijections

$$\begin{aligned} \varphi &: E \to E', \quad (x,y) \mapsto (u^{-2}(x - r), u^{-3}(y - sx - t + rs)) \quad \text{and} \\ \varphi' &: E' \to E, \quad (x,y) \mapsto (u^2x + r, u^3y + u^2sx + t) \end{aligned}$$

with $\varphi' \circ \varphi = \mathrm{id}|_E$ and $\varphi \circ \varphi' = \mathrm{id}|_{E'}$.

Under the assumption that an affine coordinate transformation is required, which assures the existence of an inverse (affine) transformation, it is easy to see that the admissible variable changes are the only possibilities. So let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \mapsto \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} r \\ t \end{pmatrix}$$

transform E to a multiple $E(\alpha X + \beta Y + r, \gamma X + \delta Y + t)$ of E'. As the degree of X in E is 3, but the degree of Y in E' only 2, we must have $\beta = 0$. The coefficients of X^3 and Y^2 in the resulting equation must be identical and not zero, i.e. $\alpha^3 = \delta^2 \neq 0$, so $\alpha = u^2$ and $\delta = u^3$ for some $u \in K^\times$. (Here it is essential that K is algebraically closed.) Finally let $s = \frac{\gamma}{u^2}$.

The coefficients a_i' of E' and the derived numbers of Definition 2.7 can be related to the corresponding values for E by plugging in the change of variables and comparing coefficients. The results are summarised in Table 2.1. Notice that isomorphic curves have the same j-invariant, which explains the name. In fact, it is not difficult to show the converse for algebraically closed K: If two curves over K have the same j-invariant, then they are isomorphic (see [Silverman, 1986], Propositions 1.4(b) and A.1.2(b)).

It is trivial to see that an admissible change of variables ψ extends to an isomorphism of $K[E]$ with $K[E']$ and consequently of $K(E)$ with $K(E')$, which we denote again by ψ. As the values of polynomial and rational functions are preserved as well, $\mathcal{O}_P(E)$ is also isomorphic via ψ to $\mathcal{O}_{\varphi(P)}(E')$ for all $P \in E$.

An important observation is that $\psi(K(X)) \subseteq K(X)$ and $\psi^{-1}(K(X)) \subseteq K(X)$, so ψ is also an automorphism of $K(X)$. This can be used to prove that ψ preserves the conjugation automorphism, formally $\psi(\bar{f}) = \overline{\psi(f)}$ for $f \in K(E)$. Let ι be the conjugation on $K(E)$. Clearly $\iota' := \psi \circ \iota \circ \psi^{-1}$ is an

$$\begin{aligned}
a'_1 &= u^{-1}(a_1 + 2s)\\
a'_3 &= u^{-3}(a_3 + ra_1 + 2t)\\
a'_2 &= u^{-2}(a_2 - sa_1 + 3r - s^2)\\
a'_4 &= u^{-4}(a_4 + 2ra_2 - (rs + t)a_1 - sa_3 + 3r^2 - 2st)\\
a'_6 &= u^{-6}(a_6 + r^2 a_2 + ra_4 - rta_1 - ta_3 + r^3 - t^2)\\
b'_2 &= u^{-2}(b_2 + 12r)\\
b'_4 &= u^{-4}(b_4 + rb_2 + 6r^2)\\
b'_6 &= u^{-6}(b_6 + 2rb_4 + r^2 b_2 + 4r^3)\\
b'_8 &= u^{-8}(b_8 + 3rb_6 + 3r^2 b_4 + r^3 b_2 + 3r^4)\\
c'_4 &= u^{-4}c_4\\
\Delta' &= u^{-12}\Delta\\
j' &= j
\end{aligned}$$

Table 2.1. Coefficients of isomorphic elliptic curves

automorphism of $K(E')$; it fixes exactly $K(X)$:

$$\begin{aligned}
& (\psi \circ \iota \circ \psi^{-1})(f) = f\\
\Leftrightarrow\quad & \iota(\psi^{-1}(f)) = \psi^{-1}(f)\\
\Leftrightarrow\quad & \psi^{-1}(f) \in K(X) \quad \text{since } \iota \text{ fixes exactly } K(X)\\
\Leftrightarrow\quad & f \in \psi(K(X)) = K(X).
\end{aligned}$$

Hence ι' is the conjugation on $K(E')$. It follows that

$$\overline{\psi(f)} = (\iota' \circ \psi)(f) = (\psi \circ \iota)(f) = \psi(\bar{f})$$

As an immediate consequence $\mathrm{N}(\psi(f)) = \psi(\mathrm{N}(f))$ and $\mathrm{Tr}(\psi(f)) = \psi(\mathrm{Tr}(f))$.

So whenever one studies rational functions or "local" properties of E, i.e. properties which can be expressed in terms of the local rings $\mathcal{O}_P(E)$, or has to deal with conjugations, traces and norms, it is sufficient to look at the isomorphism class containing E. The rest of this section is devoted to finding for each elliptic curve an isomorphic representative with a possibly "simpler" form.

If $\mathrm{char}(K) \neq 2$, the first step is completing the square on the left hand side of the Weierstraß equation. The corresponding admissible change of variables $(X, Y) \mapsto (X, Y - \frac{1}{2}(a_1 X + a_3))$ transforms E to an equation

$$E' : Y^2 = X^3 + a'_2 X^2 + a'_4 X + a'_6.$$

If furthermore $\mathrm{char}(K) \neq 3$, a similar process can be applied to the right hand side for eliminating the X^2 term. The transformation $(X, Y) \mapsto (X - \frac{1}{3}a'_2, Y)$

yields

$$E'' : Y^2 = X^3 + a_4''X + a_6''.$$

If $\mathrm{char}(K) = 3$, we want to eliminate at least one of the terms in equation E'. If $a_2' = 0$ (i.e. $j' = \frac{a_2'^6}{\Delta} = 0$ for $\Delta \neq 0$), E' is already the desired normal form. Otherwise the substitution $(X, Y) \mapsto (X + \frac{a_4'}{a_2'}, Y)$ yields a curve of the form

$$E'' : Y^2 = X^3 + a_2''X^2 + a_6''.$$

For $\mathrm{char}(K) = 2$, a similar case distinction starting with E is necessary. If $a_1 = 0$, which means $j = \frac{a_1^{12}}{\Delta} = 0$ for $\Delta \neq 0$, the substitution $(X, Y) \mapsto (X + a_2, Y)$ additionally eliminates the X^2 term. Otherwise the admissible change of variables $(X, Y) \mapsto \left(a_1^2 X + \frac{a_3}{a_1}, a_1^3 Y + \frac{a_1^2 a_4 + a_3^2}{a_1^3}\right)$ results in a curve of the form

$$E'' : Y^2 + XY = X^3 + a_2''X^2 + a_6''.$$

Table 2.2 summarises the normal forms when j is defined, i.e. $\Delta \neq 0$. As we have seen the same curves appear for $\Delta = 0$ with the appropriate case distinctions on a_2 resp. a_1.

normal form	Δ	j
$\mathrm{char}(K) \neq 2, 3$		
$Y^2 = X^3 + a_4X + a_6$	$-16(4a_4^3 + 27a_6^2)$	$1728\frac{4a_4^3}{4a_4^3+27a_6^2}$
$\mathrm{char}(K) = 3$, $j \neq 0$		
$Y^2 = X^3 + a_2X^2 + a_6$	$-a_2^3 a_6$	$-\frac{a_2^3}{a_6}$
$\mathrm{char}(K) = 3$, $j = 0$		
$Y^2 = X^3 + a_4X + a_6$	$-a_4^3$	0
$\mathrm{char}(K) = 2$, $j \neq 0$		
$Y^2 + XY = X^3 + a_2X^2 + a_6$	a_6	$\frac{1}{a_6}$
$\mathrm{char}(K) = 2$, $j = 0$		
$Y^2 + a_3Y = X^3 + a_4X + a_6$	a_3^4	0

Table 2.2. Normal forms for elliptic curves

2.4 SINGULARITIES

For dealing with singularities a simple criterion when a curve defined by a Weierstraß equation is singular would be useful. Such a criterion is provided by the discriminant.

Theorem 2.10 *A curve defined by a Weierstraß equation is singular if and only if its discriminant is zero.*

Proof: Since admissible changes of variables change Δ only by a non-zero constant (see Table 2.1) and — as affine transformations — preserve singularities, it is sufficient to prove the theorem for the normal forms shown in Table 2.2.

Consider first the case where $\operatorname{char} K \neq 2$. Then E can be supposed to be of the form $Y^2 = X^3 + a_2X^2 + a_4X + a_6$, and Δ equals 16 times the discriminant of $f = X^3 + a_2X^2 + a_4X + a_6$. So $\Delta = 0$ if and only if f has a double root. Now $P = (x, y)$ is singular on E exactly if $E(x, y) = 2y = f'(x) = 0$, or equivalently $y = 0, f(x) = f'(x) = 0$, which proves the assertion. There are two cases left for $\operatorname{char} K = 2$:

1st case: $E = Y^2 + XY + X^3 + a_2X^2 + a_6$, $\Delta = a_6$

$$\begin{aligned}\frac{\partial E}{\partial X}(X,Y) &= Y + X^2 \\ \frac{\partial E}{\partial Y}(X,Y) &= X\end{aligned}$$

So a point $P = (x, y)$ is singular on E exactly for $x = 0$, $y + x^2 = 0$ and $E(x, y) = 0$, which means that $x = y = E(0, 0) = 0$. This happens exactly for $a_6 = 0$.

2nd case: $E = Y^2 + a_3Y + X^3 + a_4X + a_6$, $\Delta = a_3^4$

$$\begin{aligned}\frac{\partial E}{\partial X}(X,Y) &= X^2 + a_4 \\ \frac{\partial E}{\partial Y}(X,Y) &= a_3\end{aligned}$$

Obviously E can only be singular for $a_3 = 0$, in which case the point $(\sqrt{a_4}, \sqrt{a_6})$ is indeed a singular point.

□

2.5 THE LOCAL RINGS $\mathcal{O}_P(E)$

We already pointed out after Definition 2.6 that $\mathcal{O}_P(E)$, the local ring of E at P, is indeed a local ring, i.e. has a unique maximal ideal $\mathfrak{m}_P$. In fact, $\mathcal{O}_P(E)$ is even a *discrete valuation ring*:

Theorem 2.11 *For an elliptic curve E and a point $P \in E$ the ring $\mathcal{O}_P(E)$ is a discrete valuation ring, i.e. there is $u \in \mathfrak{m}_P(E)$ such that each $s \in \mathcal{O}_P(E) \setminus \{0\}$ has a representation*

$$s = u^d r$$

with a non-negative integer d and a unit $r \in \mathcal{O}_P(E)^\times$.

u is called a *uniformising parameter* for $\mathcal{O}_P(E)$, and it is not difficult to see that the number d does not depend on the choice of u: In fact, u is a generator of $\mathfrak{m}_P(E)$, and d is the unique non-negative integer such that $s \in \mathfrak{m}_P(E)^d$ and $s \notin \mathfrak{m}_P(E)^{d+1}$.

Theorem 2.11 is in fact a characterisation of non-singular points on an arbitrary curve. We will establish in Section 2.9 that any line which is not tangent to the curve at the given point is a uniformising parameter. For the time being, we prove the theorem by taking specific uniformising parameters for each point. But first we have a closer look at some points which must be handled separately:

Definition 2.12 *A point $P = (x, y)$ is* of order 2 *if $P = \overline{P}$, i.e. $Y(P) = y = -y - a_1 x - a_3 = \overline{Y}(P)$.*

These points are the points of order 2 for the group law to be defined in Section 2.11, therefore the name. Let us have a closer look at the existence of such points; since isomorphisms of elliptic curves preserve the conjugation, (see Section 2.3) we consider only elliptic curves in the normal forms of Table 2.2.

char $K \neq 2$; $Y^2 = X^3 + a_2X^2 + a_4X + a_6$
(x, y) is of order 2 for $y = -y = 0$; there are exactly three such points where x is one of the (distinct) roots of $X^3 + a_2X^2 + a_4X + a_6$.

char $K = 2$, $j \neq 0$; $Y^2 + XY = X^3 + a_2X^2 + a_6$, $a_6 \neq 0$
(x, y) is of order 2 for $0 = 2y = x$. There is exactly one such point, namely $(0, \sqrt{a_6})$.

char $K = 2$, $j = 0$; $Y^2 + a_3X = X^3 + a_4X + a_6$, $a_3 \neq 0$
(x, y) is of order 2 for $0 = 2y = a_3$, so there is no such point.

Proof of Theorem 2.11: Recall that $\mathfrak{m}_P$ is the ideal formed by the functions which have a zero at P, and that the units $\mathcal{O}_P^\times$ are the set of functions regular and not zero at P.

Let $s \in \mathcal{O}_P(E)$. Then $s = \frac{f}{g}$ with f, $g \in K[E]$ and $g(P) \neq 0$. Hence g is a unit, and it is sufficient to prove the theorem for polynomial functions f only. If $f(P) \neq 0$, then f is a unit itself and $d = 0$, so we can assume that $f(P) = 0$.

- If $P = (x, y)$ is not of order 2, then $u = X - x$ is a uniformising parameter. We write $f = v + Yw$ with v, $w \in K[X]$ and split factors $X - x$ from v and w until at least one of them is no more divisible by $X - x$. Then

$f = (X - x)^{d_1}(v_1(X) + Yw_1(X))$ with $v_1(x) \neq 0$ or $w_1(x) \neq 0$. Let $f_1 = v_1 + Yw_1$. If $f_1(P) \neq 0$, then f_1 is a unit and $d = d_1$. If $\overline{f_1}(P) \neq 0$, then $\overline{f_1}$ is a unit, and $f_1 = \mathrm{N}(f_1)\overline{f_1}^{-1} = (X - x)^{d_2} f_2 \overline{f_1}^{-1}$ with $f_2 \in K[X]$ and $f_2(x) \neq 0$, so that $f_2\overline{f_1}^{-1}$ is a unit and $d = d_1 + d_2$. Let us now assume that $f_1(P) = \overline{f_1}(P) = 0$. Then $(\alpha, \beta) = (v_1(x), w_1(x))$ is a solution of the homogeneous system of linear equations

$$\begin{pmatrix} 1 & Y(P) \\ 1 & \overline{Y}(P) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 0.$$

But this system has the unique solution $\alpha = \beta = 0$ since the matrix determinant is $(\overline{Y} - Y)(P) \neq 0$ for P not of order 2. This is a contradiction to the assumption above that not both of $v_1(x)$ and $w_1(x)$ are zero.

- If P is of order 2 and $\operatorname{char} K \neq 2$, then $u = Y + \frac{1}{2}(a_1X + a_3) = \frac{1}{2}(Y - \overline{Y})$ is a uniformising parameter. By applying the change of variables $(X, Y) \mapsto (X, Y - \frac{1}{2}(a_1X + a_3))$ to the general Weierstraß equation, u and P it suffices to consider the case where

$$E : Y^2 = X^3 + a_2X^2 + a_4X + a_6, \quad u = Y, \quad P = (x_1, 0)$$

and the right hand side of E has three distinct zeros x_1, x_2 and x_3. Notice that

$$X - x_1 = \frac{(X - x_1)(X - x_2)(X - x_3)}{(X - x_2)(X - x_3)} = \frac{Y^2}{(X - x_2)(X - x_3)},$$

where the denominator is not zero in P. Write

$$f = (X - x_1)^{d_1} f_1 = \frac{Y^{2d_1}}{(X - x_2)^{d_1}(X - x_3)^{d_1}} f_1$$

with $f_1 = v_1 + Yw_1$, $v_1, w_1 \in K[X]$ and not both of $v_1(x_1)$ and $w_1(x_1)$ are zero. If $f_1(P) \neq 0$, then $d = 2d_1$. Otherwise $v_1(x_1) = 0$, so $v_1 = (X - x_1)v_2$ with $v_2 \in K[X]$ and $w_1(x_1) \neq 0$. Then

$$f_1 = \frac{(X - x_1)(X - x_2)(X - x_3)v_2 + Yw_2}{(X - x_2)(X - x_3)} = Y\frac{v_2Y + w_2}{(X - x_2)(X - x_3)},$$

where $w_2 = w_1(X - x_2)(X - x_3)$, and the second factor is a unit. In this case $d = 2d_1 + 1$.

- If $P = (x, y)$ is of order 2 and $\operatorname{char} K = 2$, then $j \neq 0$. A uniformising parameter is given by $Y + y$. Apply the change of variables $(X, Y) \mapsto (a_1^2X + \frac{a_3}{a_1}, a_1^3Y + t)$ with $t = \frac{a_1^2a_4 + a_3^2}{a_1^3}$, which transforms E to $E' : Y^2 + XY = X^3 + a_2'X^2 + a_6'$, P to (x', y') with $y' = a_1^{-3}(y + t)$ and u to $a_1^3Y + t + y =$

$a_1^3(Y+y')$. So we can assume E to be of the form E', $P=(0,y)$ with $y^2=a_6\neq 0$ and $u=Y+y$. Similarly to the previous case we compute

$$\begin{aligned} X &= (Y+y)^2\frac{X}{(Y+y)^2}=(Y+y)^2\frac{X}{Y^2+a_6} \\ &= (Y+y)^2\frac{X}{X^3+a_2X^2+XY}=\frac{(Y+y)^2}{X^2+a_2X+Y}, \end{aligned}$$

where the denominator is not zero at P. Write

$$f=X^{d_1}f_1=\frac{(Y+y)^{2d_1}}{(X^2+a_2X+Y)^{2d_1}}f_1$$

with $f_1=v_1+(Y+y)w_1$, v_1, $w_1\in K[X]$ and not both of $v_1(0)$ and $w_1(0)$ are zero. If $f_1(P)\neq 0$, then $d=2d_1$. Otherwise $v_1=Xv_2$ with $v_2\in K[X]$ and $w_1(0)\neq 0$, and

$$f_1=(Y+y)\frac{(Y+y)v_2+w_1(X^2+a_2X+Y)}{X^2+a_2X+Y},$$

where the second factor is a unit. Hence $d=2d_1+1$.

□

As usual a discrete valuation ring defines a discrete valuation (or order function) on its field of fractions: Denote for a polynomial $f\neq 0$ by $\operatorname{ord}_P(f)$ the integer d of Theorem 2.11. By convention, $\operatorname{ord}_P(0)=\infty$. The function ord_P is multiplicative from $K[E]$ to $\mathbb{Z}$ and can be extended canonically to $K(E)$ by setting $\operatorname{ord}_P\left(\frac{f}{g}\right)=\operatorname{ord}_P(f)-\operatorname{ord}_P(g)$.

Definition 2.13 *For $r\in K(E)$ and $P\in E$ the value* $\operatorname{ord}_P(r)$ *is called the* order of r at P. *If* $\operatorname{ord}_P(r)>0$ *then r is said to have a* zero *at P, if* $\operatorname{ord}_P(r)<0$ *then r is said to have a* pole *at P. The* multiplicity *of the zero or pole is* $|\operatorname{ord}_P(r)|$.

Notice that the proof of Theorem 2.11 is constructive and allows to determine the order of a rational function at any given point.

Example. Let $P=(x,y)\in E$. We want to determine all zeros of $X-x$ and their orders.

- If P is not of order 2, then there are exactly two points with X-coordinate x, namely P and $\overline{P}$. $X-x$ is a uniformising parameter for both of the local rings, so $X-x$ has simple zeros at P and $\overline{P}$ and order zero at any other point of E.
- Now we assume that char $K\neq 2$, P is of order 2 and E is in normal form. Then $P=(x_1,0)$ and $E: Y^2=(X-x_1)(X-x_2)(X-x_3)$ with x_1,x_2 and x_3 distinct. As Y is a uniformising parameter and $(X-x_2)(X-x_3)$ is not

zero at P, the order of $X - x_1$ is two. By applying the change of variables of Theorem 2.11 in the backward direction one easily sees that this is still true if E is not in normal form: $X - x$ has a double zero at P and order zero at any other point.

- If $\operatorname{char} K = 2$, P is of order 2 and E in normal form $Y^2 + XY = X^3 + a_2X^2 + a_6$, then $x = 0$ and $y^2 = a_6 \neq 0$. There is only the point P with X-coordinate zero; its uniformising parameter is $Y + y$. As we have seen in the proof of Theorem 2.11

  $$X = (Y + y)^2 \frac{1}{X^2 + a_2X + Y}$$

  and

  $$\frac{1}{X^2 + a_2X + Y}(P) = \frac{1}{y}$$

  is defined and not zero. So X has a double zero at P and order zero at any other point. Again the change of variables needed to obtain the normal form can be made explicit, and the result stays correct if E is in general form.

We recall a basic fact about discrete valuations which will be useful for later proofs:

Proposition 2.14 *Let* ord *be a discrete valuation on a field L. Then*

$$\operatorname{ord}(f + g) \geq \min\{\operatorname{ord} f, \operatorname{ord} g\} \quad \forall f, g \in L$$

with equality if $\operatorname{ord} f \neq \operatorname{ord} g$. *More generally*

$$\operatorname{ord}\left(\sum_{i=1}^{n} f_i\right) \geq \min\{\operatorname{ord} f_i : 1 \leq i \leq n\} \quad \forall n \in \mathbb{N},\ f_1, \ldots, f_n \in L$$

with equality if the minimum is unique.

Proof: The assertion is an immediate consequence of Theorem 2.11 in our setting. In a more general context one either defines first discrete valuation rings as in Theorem 2.11; then the proposition follows in the same way. Or the proposition itself is part of the definition of a discrete valuation, and the notion of a discrete valuation ring is derived from it. □

2.6 PROJECTIVE PLANE CURVES

For constructing the group law on an elliptic curve we need that any line intersects the curve in three points; the sum of two points is more or less the third point of intersection of the line through these points with the curve. More generally spoken, we make use of Bézout's famous theorem that two distinct curves of degree m and n, respectively, intersect in mn points, counting multiplicities in an appropriate way: “Le degré de l’équation finale résultante d’un nombre quelconque d’équations complettes renfermant un pareil nombre d’inconnues, & de degrés quelconques, est égal au produit des exposans des degrés de ces équations.” ([Bézout, 1779], p. 32). Obviously this theorem is not true in our actual affine setting: Two parallel lines do not even intersect at all. So we have to add extra points; it turns out that it is sufficient to add one point for each parallel class of lines to make Bézout’s theorem work. The resulting object is called the *projective plane.* We then have to specify which of the newly added points are added to a given curve, thus constructing a *projective curve* from an affine one.

Definition 2.15 *The* projective plane *over K is the set of equivalence classes of $K^3\setminus\{(0,0,0)\}$ under the equivalence relation*

$$(x_1,y_1,z_1) \sim (x_2,y_2,z_2) \Leftrightarrow \exists\lambda \in K^\times : (x_2,y_2,z_2) = \lambda(x_1,y_1,z_1)$$

It is denoted by $P^2(K)$.

More intuitively spoken we take as “points” the lines through the origin in the three dimensional affine space. From now on (x,y,z) denotes the whole equivalence class $K^\times(x,y,z)$.

Curves are now defined as polynomials in $K[X,Y,Z]$; but we must make sure that a polynomial which has a zero in a representative (x,y,z) actually has a zero in all elements of the equivalence class $K^\times(x,y,z)$. This is achieved by admitting only homogeneous polynomials, i.e. polynomials each monomial of which has the same degree. The set of homogeneous polynomials is denoted by $K[X,Y,Z]_{\text{hom}}$.

Definition 2.16 *A* projective plane curve *is the set of zeros in the projective plane of an irreducible homogeneous polynomial C in $K[X,Y,Z]_{\text{hom}}$. A point P on C is* singular *if*

$$\frac{\partial C}{\partial X}(P) = \frac{\partial C}{\partial Y}(P) = \frac{\partial C}{\partial Z}(P) = 0$$

Until now it is not clear at all why we have “added points” to the affine plane or to an affine curve; we have to identify A^2 with a certain subset of P^2 and define corresponding operations on polynomials such that an affine point lies on an affine curve exactly if its projective representative lies on the transformed projective curve. We start with the points. Let U be the subset of P^2 consisting of points with Z-coordinate not zero, called the *finite points* of P^2. Then — by dividing through the Z-coordinate — any point in U has a unique representative $(x,y,1)$. Such a point is identified with the affine point (x,y).

Definition 2.17 *The maps*

$$A^2 \to U, \qquad (x,y) \mapsto (x,y)^* = (x,y,1)$$

and

$$U \to A^2, \qquad (x,y,z) \mapsto (x,y,z)_* = \left(\frac{x}{z}, \frac{y}{z}, 1\right)_* = \left(\frac{x}{z}, \frac{y}{z}\right)$$

are called homogenisation *and* dehomogenisation *of points with respect to Z. They are inverse bijections of A^2 with U.*

The points with Z-coordinate zero are called *points at infinity* because they would require a division by zero. They are the "new" points in the projective plane. Our aim is now to define homogenisation and dehomogenisation maps between $K[X,Y]$ and $K[X,Y,Z]_{\mathrm{hom}}$ in such a way that $P \in A^2(K)$ is a zero of $f \in K[X,Y]$ if and only if P^* is a zero of f^*, and $P \in U$ is a zero of $f \in K[X,Y,Z]_{\mathrm{hom}}$ if and only if P_* is a zero of f_*. The second operation is easier to define: For the equivalence $f(x,y,1) = 0 \Leftrightarrow f_*(x,y) = 0$ it suffices to let $f_* = f(X,Y,1)$. In the other direction we require $f\left(\frac{x}{z}, \frac{y}{z}\right) = 0 \Leftrightarrow f^*(x,y,z) = 0$. As $f^* = f\left(\frac{x}{z}, \frac{y}{z}\right)$ does not define a polynomial, one has to multiply by an appropriate power of Z, letting $f^* = Z^{\deg f} f\left(\frac{x}{z}, \frac{y}{z}\right)$.

Definition 2.18 *The maps*

$$K[X,Y] \to K[X,Y,Z]_{\mathrm{hom}}, \qquad f \mapsto f^* = Z^{\deg f} f\left(\frac{x}{z}, \frac{y}{z}\right)$$

and

$$K[X,Y,Z]_{\mathrm{hom}} \to K[X,Y], \qquad f \mapsto f_* = f(X,Y,1)$$

are called homogenisation *and* dehomogenisation *with respect to Z.*

We still have to check that if C is an affine curve, then C^* is a projective one and vice versa, i.e. that irreducibility is preserved by homogenisation and dehomogenisation. This is a corollary of the following proposition whose easy proof is left as an exercise to the reader.

Proposition 2.19 *Let $f, g \in K[X,Y]$, $F, G \in K[X,Y,Z]_{\mathrm{hom}}$.*

1. $(fg)^* = f^* g^*$

2. $(FG)_* = F_* G_*$

3. $(f^*)_* = f$

4. *If Z does not divide F, then $(F_*)^* = F$.*

Corollary 2.20 *A polynomial $C \in K[X,Y]$ defines a curve if and only if C^* does.*

Definition 2.21 *If C is an affine curve then C^* is called its* projective closure. *It consists of the points of C and possibly additional points at infinity.*

As in Section 2.1 we want to define rational functions on C^*. To be well defined, such a function r must be homogeneous of degree zero:

$$r(\lambda P) = \lambda^0 r(P) = r(P) \quad \forall \lambda \in K^\times, P \in C^*$$

Definition 2.22 *The* field of rational functions on C^**, denoted by* $K(C^*)$*, is the subfield of the field of fractions of* $K[C^*] := K[X,Y,Z]/(C^*)$ *consisting of the zero function and quotients of the form* $r = \frac{f}{g}$ *where* f *and* g *are homogeneous of the same degree. A rational function* r *is* regular *or* defined *at* $P = (x,y,z) \in C^*$ *if it has a representation* $r = \frac{f}{g}$ *where* f *and* g *are homogeneous of the same degree, with* $g(P) \neq 0$*. The ring of all such functions is called the* local ring of C^* at P *and denoted by* $\mathcal{O}_P(C^*)$*.*

It is easy to verify that the so defined sets are indeed fields and local rings. We would now like to relate the fields of rational functions of an affine curve and of its projective closure by extending the homogenisation and dehomogenisation maps. This can be done without problem for the dehomogenisation map: If $r = \frac{f}{g} \in K(C^*)$ with f, $g \in K[C^*]$, we set $r_* = \frac{f_*}{g_*} = \frac{f(X,Y,1)}{g(X,Y,1)}$. For $r = \frac{f}{g} \in K(C)$ with f, $g \in K[C]$, however, the obvious definition $r^* = \frac{f^*}{g^*}$ does not have the desired result. Remember that $f^* = Z^{\deg f} f\left(\frac{X}{Z}, \frac{Y}{Z}\right)$ has the same degree as f; so if $\deg f \neq \deg g$, f^* and g^* are homogeneous, but not of the same degree, and $\frac{f^*}{g^*} \notin K(C^*)$. Again we multiply by just the power of Z which solves the problem and set $r^* = \frac{f(X/Z,Y/Z)}{g(X/Z,Y/Z)} = Z^{\deg g - \deg f} \frac{f^*}{g^*}$.

This leaves us with an ambiguity because for $f \in K[C]$, two homogenised forms are possible: On one hand, f can be seen as a polynomial, and $f^* = Z^{\deg f} f\left(\frac{X}{Z}, \frac{Y}{Z}\right)$ is the corresponding homogenised polynomial. On the other hand, $f = \frac{f}{1}$ is a rational function, and $f^* = f\left(\frac{X}{Z}, \frac{Y}{Z}\right)$ is the corresponding homogenised rational function. Depending on the property we would like to preserve, we use any of the homogenised forms, arguing that powers of Z have no effect when dehomogenising again.

Proposition 2.23 *The homogenisation and dehomogenisation maps defined by*

$$K(C) \to K(C^*), \qquad \frac{f}{g} \mapsto Z^{\deg g - \deg f} \frac{f^*}{g^*} = \frac{f(\frac{x}{z}, \frac{y}{z})}{g(\frac{x}{z}, \frac{y}{z})}$$

and

$$K(C^*) \to K(C), \qquad \frac{f}{g} \mapsto \frac{f_*}{g_*} = \frac{f(X,Y,1)}{g(X,Y,1)}$$

are inverse field isomorphisms. If P *is a point of* C*, then their restrictions to* $\mathcal{O}_P(C)$ *and* $\mathcal{O}_{P^*}(C^*)$ *are inverse isomorphisms of local rings, and* $r(P) = r^*(P^*)$ *for all* $r \in \mathcal{O}_P(C)$*.*

Proof: For the first part it is trivial to verify that the operations are additive and multiplicative homomorphisms and that one is the inverse of the other. For the second part note that $f(x,y) = f^*(x,y,1)$. □

We defined homogenisation and dehomogenisation with respect to the Z-coordinate; of course the same procedure can be applied to X and Y, which allows to handle infinite points. Since at least one coordinate of any point is not zero, it can be regarded as finite when dehomogenising with respect to this coordinate. So all points can be worked with "affinely" if desired. In the following sections we will switch deliberately between the affine and the projective point of view.

2.7 PROJECTIVE ELLIPTIC CURVES

A projective elliptic curve is defined in the obvious way as the projective closure of an affine one. Then many of the results from the previous sections generalise without difficulty to the projective case.

Definition 2.24 *A* projective Weierstraß equation *is an equation of the form*

$$E : Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3.$$

Its discriminant Δ and its j-invariant are as in Definition 2.7. A non-singular projective Weierstraß equation defines a projective elliptic curve. *Two curves defined by projective Weierstraß equations*

$$\begin{aligned} E &: Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3 \\ E' &: Y^2Z + a_1'XYZ + a_3'YZ^2 = X^3 + a_2'X^2Z + a_4'XZ^2 + a_6'Z^3 \end{aligned}$$

are isomorphic *if E' can be obtained from E by a change of variables of the form*

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \mapsto \begin{pmatrix} u^2X + rZ \\ u^3Y + u^2sX + tZ \\ Z \end{pmatrix}$$

where $u \in K^\times$, r, s, $t \in K$ (and dividing the resulting equation by u^6). The corresponding transformation is referred to as an admissible change of variables.

Proposition 2.25 *Any projective Weierstraß equation is irreducible. It contains the unique infinite point $\mathcal{O} = (0, 1, 0)$, and is singular if and only if $\Delta = 0$.*

Proof: The irreducibility follows directly from the irreducibility of E_* (see the discussion after Definition 2.7) and Proposition 2.19. Let $P = (x, y, z) \in E$ be an infinite point, i.e. $z = 0$. Then $x^3 = 0$, and $P = \mathcal{O}$. For the last part use Theorem 2.10 and verify that a Weierstraß equation is never singular at $\mathcal{O}$:

$$\frac{\partial E}{\partial Z}(0, 1, 0) = 1 \neq 0$$

□

Theorem 2.26 *For a projective elliptic curve E and a point $P \in E$ the ring $\mathcal{O}_P(E)$ is a discrete valuation ring.*

Proof: The case where P is finite is exactly Theorem 2.11. For $P = \mathcal{O}$ a uniformising parameter is given by $u = \frac{X}{Y}$. Dehomogenise the Weierstraß equation with respect to Y:

$$E_* : Z + a_1XZ + a_3Z^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3.$$

In the light of Proposition 2.23 it must now be shown that $u_* = X$ is a uniformising parameter for $\mathcal{O}_{(0,0)}(E_*)$, where $(0,0) = \mathcal{O}_*$. Notice first that X divides Z in $\mathcal{O}_{(0,0)}(E_*)$:

$$\begin{aligned} Z &= \frac{ZX^3}{X^3} \\ &= \frac{ZX^3}{Z + a_1XZ + a_3Z^2 - a_2X^2Z - a_4XZ^2 - a_6Z^3} \\ &= X^3\frac{1}{1 + a_1X + a_3Z - a_2X^2 - a_4XZ - a_6Z^2}, \end{aligned}$$

where the denominator is not zero at $(0,0)$. Now for a polynomial $f \in K[E_*]$, write $f = r(Z) + s(Z)X + t(Z)X^2$ with $r, s, t \in K[Z]$ by repeated replacing of X^3 with the corresponding term of E_*. Then write $f = r_1(Z)Z^i + s_1(Z)Z^jX + t_1(Z)Z^kX^2$ with each of r_1, s_1 and t_1 being either zero or not divisible by Z. Replace Z by

$$\frac{X^3}{1 + a_1X + a_3Z - a_2X^2 - a_4XZ - a_6Z^2};$$

then $f = r_2(X,Z)X^{3i} + s_2(X,Z)X^{3j+1} + t_2(X,Z)X^{3k+2}$ where r_2, s_2 and t_2 are rational functions which are regular and either the zero polynomial or not zero at $(0,0)$. Let d be the minimum of the numbers $3i$ (if $r_2 \neq 0$), $3j+1$ (if $s_2 \neq 0$) and $3k+2$ (if $t_2 \neq 0$). Then $f = X^df'$ with f' regular and not zero at $(0,0)$ because d equals exactly one of the numbers $3i$, $3j+1$ and $3k+2$. □

Example. We have just computed that $\frac{Z}{Y}$ has a zero of order 3 at $\mathcal{O}$ on E (or affinely spoken Z has a zero of order 3 at $(0,0)$ on E_* where dehomogenisation is done with respect to Y). Similarly, $\frac{Y}{Z}$ or $\frac{1}{Z}$ has a pole of order 3. As $\text{ord}_{\mathcal{O}}\left(\frac{X}{Y}\right) = 1$ and $\text{ord}_{\mathcal{O}}\left(\frac{Z}{Y}\right) = 3$, the rational function $\frac{X-xZ}{Y}$ has a simple zero at $\mathcal{O}$ (see Proposition 2.14) and thus $\frac{X-xZ}{Z} = \frac{X-xZ}{Y}\frac{Y}{Z}$ has a pole of order 2 at $\mathcal{O}$ for $x \in K$. This observation completes the determination of the zeros of $X - x$ in the example after Definition 2.13.

2.8 DIVISORS

To keep track of the zeros and poles of a rational function it is useful to code their multiplicities into a new object, namely as coefficients in the free abelian group generated by the points of E.

Definition 2.27 *The* group of divisors *of E is the free abelian group generated by the points of E,*

$$\operatorname{Div}(E) = \left\{ \sum_{P \in E} m_P \langle P \rangle : m_P \in \mathbb{Z}, m_P = 0 \textit{ for all but finitely many } P \in E \right\};$$

the sums have to be seen as purely formal and are not to be confused with the addition on elliptic curves, which will be defined in Section 2.10. If $\Delta = \sum_{P \in E} m_P \langle P \rangle$ is a divisor, then its degree *is* $\deg \Delta = \sum_{P \in E} m_P$. *The subgroup of* $\operatorname{Div}(E)$ *of divisors of degree zero is denoted by* $\operatorname{Div}^0(E)$. *To a rational function $r \neq 0$ the divisor* $\operatorname{div} r = \sum_{P \in E} \operatorname{ord}_P(r) \langle P \rangle$ *is assigned. Such a divisor is called* principal.

We will show in Corollary 2.30 that a rational function has only a finite number of zeros and poles, so $\operatorname{div} r$ is indeed a divisor.

Example. Let $P = (x, y, 1) \in E$. Summarising the examples of the Sections 2.5 and 2.7,

$$\operatorname{div}\left(\frac{X - xZ}{Z}\right) = \langle P \rangle + \langle \overline{P} \rangle - 2\langle O \rangle.$$

Since $\operatorname{ord}_P(r_1 r_2) = \operatorname{ord}_P(r_1) + \operatorname{ord}_P(r_2)$, the map

$$\operatorname{div} : K(E)^\times \to \operatorname{Div}(E)$$

is a homomorphism of abelian groups. Hence the set of principal divisors is a subgroup of $\operatorname{Div}(E)$; it is denoted by $\operatorname{Prin}(E)$.

When $\varphi : E \to E'$ is an isomorphism of elliptic curves and $\psi : K(E) \to K(E')$ the corresponding admissible change of variables, then φ induces a group isomorphism: $\operatorname{Div}(E) \to \operatorname{Div}(E')$ by letting $\varphi(\langle P \rangle) = \langle \varphi(P) \rangle$. We observed on page 18 that ψ constitutes an isomorphism of $\mathcal{O}_P(E)$ with $\mathcal{O}_{\varphi(P)}(E)$ for any point P on E. Consequently, we have

$$\operatorname{div}(\psi(r)) = \varphi(\operatorname{div}(r)) \text{ for } r \in K(E),$$

and φ is also an isomorphism of the subgroups of principal divisors $\operatorname{Prin}(E)$ and $\operatorname{Prin}(E')$. Especially,

$$\operatorname{div}(\bar{r}) = \overline{\operatorname{div}(r)},$$

so $\bar{r}$ has the same multiplicities of zeros and poles as r (but usually at different points).

The rest of this section is devoted to counting the zeros and poles of rational functions; the main result is the following theorem:

Theorem 2.28 *A rational function has as many zeros as poles, counting multiplicities; hence* $\operatorname{Prin}(E) \subseteq \operatorname{Div}^0(E)$.

We first observe that it is enough to prove the theorem for rational functions of the form $f^* = f(\frac{X}{Z}, \frac{Y}{Z})$ with $f \in K[E_*]$, for any rational function is a quotient of these special ones (see Proposition 2.23). We then count the order of f^* separately at finite points (where it has zeros) and $\mathcal{O}$ (where it has a pole).

Lemma 2.29 *$f \in K[E_*]$ has* $\deg(\mathrm{N}(f))$ *many zeros at points of E_*, counting multiplicities; here* deg *denotes the usual degree in X.*

Proof: Let $n = \deg(\mathrm{N}(f))$, and write $f\bar{f} = \mathrm{N}(f) = (X - x_1) \cdots (X - x_n)$. We have seen in the example of Section 2.5 that $X - x_i$ has two zeros in E_*, so $f\bar{f}$ has precisely $2n$ zeros. As f and $\bar{f}$ must have the same number of zeros by the assertion above, this number is n. □

An immediate consequence of this lemma is the following observation:

Corollary 2.30 *A rational function has only a finite number of zeros and poles.*

Lemma 2.31 *If $f = v(X) + Yw(X) \in K[E_*]$ then f^* has a pole of order* $\max\{2\deg v, 2\deg w + 3\}$ *at $\mathcal{O}$.*

Note that the term in this lemma is the definition in [Charlap and Robbins, 1988] of the order of a rational function at $\mathcal{O}$; our use of the projective setting gives a further justification of this definition.

Proof: Consider first the case where $w = 0$; let $n = \deg v$ such that

$$f^* = \sum_{i=0}^{n} a_i \frac{X^i}{Z^i} \text{ with } a_n \neq 0.$$

We have to show that $\mathrm{ord}_{\mathcal{O}}(f^*) = -2n$. If $a_i \neq 0$, then $\mathrm{ord}_{\mathcal{O}}\left(a_i \frac{X^i}{Z^i}\right) = -2i$ by the example at the end of Section 2.7. These expressions have a unique minimum at $i = n$. It follows from Proposition 2.14 that $\mathrm{ord}_{\mathcal{O}}(f^*) = -2n$.

It remains to consider the case $f^* = v\left(\frac{X}{Z}\right) + \frac{Y}{Z} w\left(\frac{X}{Z}\right)$ with $w \neq 0$. Then

$$\mathrm{ord}_{\mathcal{O}}\left(v\left(\frac{X}{Z}\right)\right) = -2\deg v$$

if $v \neq 0$, following the first part of the proof, and

$$\mathrm{ord}_{\mathcal{O}}\left(\frac{Y}{Z} w\left(\frac{X}{Z}\right)\right) = \mathrm{ord}_{\mathcal{O}}\left(\frac{Y}{Z}\right) + \mathrm{ord}_{\mathcal{O}}\left(w\left(\frac{X}{Z}\right)\right) = -3 - 2\deg w$$

by the example at the end of Section 2.7 and the first part of the proof applied to w. If $v = 0$, this finishes the proof. Otherwise, $-2\deg v$ and $-3 - 2\deg w$ are distinct because one is even and one is odd, and Proposition 2.14 shows the assertion. □

The next lemma finishes the proof of Theorem 2.28.

Lemma 2.32 *If $f = v(X) + Yw(X) \in K[E_*]$, then*

$$\deg(\mathrm{N}(f)) = \max\{2\deg v, 2\deg w + 3\}.$$

Proof: By the example after Definition 2.8,

$$\mathrm{N}(f) = v^2 + \mathrm{Tr}(Y)w + \mathrm{N}(Y)w^2,$$

where

- $\deg(v^2) = 2\deg v$ is even,
- $\deg(\mathrm{N}(Y)w^2) = 3 + 2\deg w$ is odd and
- $\deg(\mathrm{Tr}(Y)w) \leq 1 + \deg w < 3 + 2\deg w$.

Then $\deg(\mathrm{N}(f)) = \max\{2\deg v, 3 + 2\deg w\}$. □

Corollary 2.33 *A non-constant polynomial has at least two finite zeros.*

Proof: The assertion follows directly from Lemmata 2.29 and 2.32. □

Proposition 2.34 *Let $r \in K(E)$ be a rational function without finite poles. Then r_* is a polynomial.*

Proof: Write $r_* = v + wY$ with $v, w \in K(X)$. By Proposition 2.14, there are two possibilities: Either v and w have no poles (during the proof, by "poles" we mean "finite poles") — then they must be polynomials, and we are finished; or they have poles in the same points. We assume the latter case and show that it does not occur; to simplify the argumentation we can assume E in one of the normal forms, since admissible changes of variables preserve polynomials and the existence of poles. As r_* has no poles, the same holds for $\overline{r}_*$ and

$$r_* - \overline{r}_* = (2Y + a_1X + a_3)w.$$

So if w has a pole in $P = (x, y)$, then P is a zero of $2Y + a_1X + a_3$ and hence a point of order 2.

1st case: $\mathrm{char}\, K \neq 2$
There are three points of order 2, and by Lemmata 2.29 and 2.32 the polynomial $2Y + a_1X + a_3$ has simple zeros in them. Thus P is a simple pole of w. On the other hand write $w = \frac{f}{g}$ with coprime polynomials $f, g \in K[X]$. Then for w to have a pole in P, it is necessary that $X - x$ divides g, but not f. Since $X - x$ has a double zero in P by the example in Section 2.5, w must have at least a double pole in P, a contradiction.

2nd case: $\mathrm{char}\, K = 2$
Then $j \neq 0$, and P is the unique point of order 2. We assume E_* in the normal form $Y^2 + XY = X^3 + a_2X^2 + a_6$ and $P = (0, \sqrt{a_6})$. Since $r_* - \overline{r}_* =$

$Xw \in K(X)$ has no poles, it is a polynomial $w_1 X + w_0$ with $w_1 \in K[X]$ and $w_0 \in K$. Furthermore $w_0 \neq 0$ because otherwise w would not have a pole in P. Then w has a double pole in P, and so must do v by Proposition 2.14, and Xv is a polynomial $v_1 X + v_0$ with $v_1 \in K[X]$ and $v_0 \in K^\times$ by the same argumentation as for w. Then

$$\begin{aligned} r_* &= \frac{v_1 X + v_0}{X} + \frac{w_1 X + w_0}{X} Y \\ &= (v_1 + w_1 Y) + w_0 \frac{Y + \frac{v_0}{w_0}}{X}. \end{aligned}$$

Since X has a double zero in P and r_* has no poles, $Y + \frac{v_0}{w_0}$ must have at least a double zero in P, which implies $\frac{v_0}{w_0} = \sqrt{a_6}$. But then the zero is simple since $Y + \sqrt{a_6}$ is a uniformising parameter for $\mathcal{O}_P(E_*)$ by the proof of Theorem 2.11, a contradiction.

□

Corollary 2.35 *Let Δ be a principal divisor. Then the rational function with divisor Δ is unique up to multiplication by a non-zero constant in K.*

Proof: Suppose $g, \tilde{g} \in K(E)$ are both rational functions with divisor Δ. Then the divisor of $r = g/\tilde{g}$ is zero. By Proposition 2.34, r_* is a polynomial, and by Corollary 2.33 it is constant. □

We are going to study which poles and zeros rational functions can have, i.e. which divisors are principal. More generally, we are interested in the quotient group of $\operatorname{Div}(E)$ by the subgroup of principal divisors. With regard to Theorem 2.28 we can restrict ourselves to divisors of degree zero. This leads to the following definition.

Definition 2.36 *Two divisors Δ_1 and Δ_2 are called* linearly equivalent *if $\Delta_1 - \Delta_2 \in \operatorname{Prin}(E)$; we write $\Delta_1 \sim \Delta_2$. $\operatorname{Pic}(E) := \operatorname{Div}(E)/\operatorname{Prin}(E)$ is called the* Picard *or* divisor class group *of E, $\operatorname{Pic}^0(E) := \operatorname{Div}^0(E)/\operatorname{Prin}(E)$ its* degree zero part.

The terminology "linearly equivalent" may seem confusing; since principal divisors are derived from rational functions, it would be more appropriate at the moment to speak of "rational equivalence". It turns out, however, that all rational functions on elliptic curves can be built from the simplest, namely the linear ones. So before continuing the general theory of divisors in Section 2.10 we have a closer look at lines. Their poles and zeros can be computed explicitly, and we will make use of this knowledge for manipulating the divisors of $\operatorname{Pic}^0(E)$.

2.9 LINES

Definition 2.37 *An (affine)* line *l is a polynomial of degree 1, i.e. of the form $l = \alpha X + \beta Y + \gamma$ with α and β not both zero. By multiplying with a factor in $K^\times$ one of α or β can be supposed to be 1.*

Interpreting lines as rational functions, two questions are of interest: First, being given a line, can its divisor be computed effectively? And second, being given a divisor or some of its entries, is there a line with this divisor?

The first question has partly been answered by the examples of the previous sections: If $l = X - x$, then $\operatorname{div}(l^*) = P + \overline{P} - 2\mathcal{O}$, where P_* is one of the points on E_* with X-coordinate x. It can be determined by plugging in x into the affine Weierstraß equation and solving the resulting quadratic equation in Y to obtain $Y(P_*)$ and $Y(\overline{P_*})$.

A similar result is true for lines with $\beta \neq 0$; in this case, β can be assumed to be 1.

Lemma 2.38 *Let $l = Y - (mX + b)$ and $P = (x, y, 1)$ a point on $l^* \cap E$. Then* $\operatorname{ord}_P(l^*)$ *is the multiplicity of x as zero of the polynomial in one variable $E_*(X, mX + b)$.*

Proof: On one hand, $E_*(X,T) = T^2 - \operatorname{Tr}(Y)T + \operatorname{N}(Y) = (Y - T)(\overline{Y} - T)$ implies

$$E_*(X, mX + b) = (Y - (mX + b))(\overline{Y} - (mX + b)) = l\overline{l};$$

on the other hand $E_*(X, mX + b)$ is a cubic polynomial and therefore splits into three linear factors:

$$E_*(X, mX + b) = -(X - x_1)(X - x_2)(X - x_3),$$

where x_1, x_2 and x_3 are not necessarily distinct. In the latter representation the divisor is known from the examples:

$$\begin{aligned} \operatorname{div}(l^*\overline{l}^*) &= \operatorname{div}(l^*) + \operatorname{div}(\overline{l}^*) \\ &= \operatorname{div}((Y - (mX + b))^*) + \operatorname{div}((\overline{Y} - (mX + b))^*) \\ &= \langle P_1 \rangle + \langle \overline{P_1} \rangle + \langle P_2 \rangle + \langle \overline{P_2} \rangle + \langle P_3 \rangle + \langle \overline{P_3} \rangle - 6\langle \mathcal{O} \rangle, \end{aligned}$$

where P_i is a zero of $(X - x_i)^*$ on E.

Denote the multiplicity of x as zero of $E_*(X, mX + b)$ by d. Then x appears d times among x_1, x_2 and x_3. By definition of P_i, we have $x = x_i$ exactly for $P \in \{P_i, \overline{P_i}\}$. Two cases can be distinguished:

1st case: $P = \overline{P}$
Then $x = x_i$ means $P = P_i = \overline{P_i}$, and P appears $2d$ times in the divisor. Thus,

$$2d = \operatorname{ord}_P(l^*) + \operatorname{ord}_P(\overline{l}^*) = \operatorname{ord}_P(l^*) + \operatorname{ord}_{\overline{P}}(l^*) = 2\operatorname{ord}_P(l^*).$$

2nd case: $P \neq \overline{P}$
Then $x = x_i$ means that P equals exactly one of P_i and $\overline{P_i}$, so

$$d = \operatorname{ord}_P(l^*) + \operatorname{ord}_P(\overline{l}^*) = \operatorname{ord}_P(l^*) + \operatorname{ord}_{\overline{P}}(l^*).$$

Moreover there is a unique line through P_* and $\overline{P_*}$, namely $X - x$. As P lies on l^*, but l is — even up to multiplication by a constant — different from $(X - x)^*$, it follows that $\overline{P} \notin l^*$ and hence $\mathrm{ord}_{\overline{P}}(l^*) = 0$.

□

The lemma suggests an easy algorithm for computing the divisor of $l = Y - (mX + b)$: Compute the roots x_1, x_2 and x_3 of $E_*(X, mX + b)$ and set $P_i = (x_i, mx_i + b, 1)$. Then $\mathrm{div}(l^*) = \langle P_1\rangle + \langle P_2\rangle + \langle P_3\rangle - 3\langle \mathcal{O}\rangle$.

Corollary 2.39 *Let l be a line. Then there are finite points P, Q and R, which can be computed effectively, such that*

$$\begin{aligned} \mathrm{div}(l^*) &= \langle P\rangle + \langle \overline{P}\rangle - 2\langle \mathcal{O}\rangle & \textit{if } l = X - x \\ \mathrm{div}(l^*) &= \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle \mathcal{O}\rangle & \textit{if } l = Y - (mX + b) \end{aligned}$$

Taking into account that the term "$-3\langle \mathcal{O}\rangle$" comes from the division by Z when homogenising, this corollary is Bézout's theorem for the intersection of lines and elliptic curves: Two curves of degree 1 and 3, respectively, intersect in three points, P, $\overline{P}$ and $\mathcal{O}$ in the first and P, Q and R in the second case.

Our second task is now to find lines with prescribed orders at certain points; this allows us to change divisors at given points by adding divisors of lines while staying in the same divisor class. As two (distinct) points already determine a line it seems natural to fix only two zeros. Formally the question is the following: Given points P and Q, not necessarily distinct, is there a line l and a point R with $\mathrm{div}(l^*) = \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle \mathcal{O}\rangle$? And if the question is answered positively, can this line be computed?

Theorem 2.40 *Let P and Q be points on E, not both equal to $\mathcal{O}$. Then there is a unique line l and a unique point R, which can be computed effectively, such that*

$$\mathrm{div}(l^*) = \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle \mathcal{O}\rangle.$$

Proof: If $P = (x, y, 1), Q = \mathcal{O}$ (or vice versa), then with Corollary 2.39, l must be $X - x$ and $R = \overline{P}$.

If $P = (x_1, y_1, 1)$ and $Q = (x_2, y_2, 1)$ are distinct finite points then there is a unique line l through P and Q. If $P = \overline{Q}$, i.e. $x_1 = x_2$, then $l = X - x_1$ and $R = \mathcal{O}$. Otherwise $l = Y - (\lambda X + \mu)$ with $\lambda = \frac{y_2 - y_1}{x_2 - x_1}$, $\mu = y_1 - \lambda x_1$, and R can be computed as above.

The remaining case $P = Q$ causes difficulties. Which line has a multiple zero at a given point? It turns out that the tangent line does, but this result needs some preparation.

Definition 2.41 *Let $P = (x, y)$ be a non-singular point on an affine curve C. Then the* tangent line at C in P *is defined to be*

$$\frac{\partial C}{\partial X}(P)(X - x) + \frac{\partial C}{\partial Y}(P)(Y - y)$$

Note that the tangent line is well defined as at least one of the partial derivatives does not vanish by definition of non-singular points.

Example. Let E be an affine elliptic curve, $P = (x, y)$. Then the tangent line at E in P is given by

$$(a_1 y - (3x^2 + 2a_2 x + a_4))(X - x) + (2y + a_1 x + a_3)(Y - y).$$

The following lemma relates the tangent through P with the order function in P. It is true for any curve, but again is proved only for elliptic curves.

Lemma 2.42 *Let P be a finite point on an elliptic curve and l a line. Then l is tangent in P if and only if* $\mathrm{ord}_P(l^*) \geq 2$.

Proof: P does not lie on l^* exactly if $\mathrm{ord}_P(l^*) = 0$. So we can assume that $P = (x, y, 1) \in l^*$, and l has the form $\alpha(X - x) + \beta(Y - y)$ with one of α and β equal to 1. As usual we have to distinguish two cases:

1st case: $P = \overline{P}$
Then $2y + a_1 x + a_3 = 0$ and the tangent line is $t = X - x$ with $\mathrm{ord}_P(t^*) = 2$. For the converse implication let us assume that $\mathrm{ord}_P(l^*) \geq 2$ and E is in one of the normal forms of Table 2.2; this is possible because admissible changes of variables are affine transformations and preserve lines and tangents. Then regardless of the characteristic, by Section 2.5, $Y - y$ is a uniformising parameter, so $\mathrm{ord}_P((Y - y)^*) = 1$, and we already computed $\mathrm{ord}_P((X - x)^*) = 2$. Hence by Proposition 2.14, $\mathrm{ord}_P(l^*) \geq 2$ is only possible if $\beta = 0$, thus $l = t$.

2nd case: $P \neq \overline{P}$
If $\beta = 0$, then l is not tangent and $\mathrm{ord}_P(l^*) = 1$ by Corollary 2.39. Assume now that $\beta = 1$. The line l is tangent exactly for

$$\alpha = \frac{\frac{\partial E_*}{\partial X}(P_*)}{\frac{\partial E_*}{\partial Y}(P_*)}.$$

On the other hand Lemma 2.38 states that $\mathrm{ord}_P(l^*) \geq 2$ if and only if $E_*(X, -\alpha(X - x) + y)$ has a multiple zero in x, or equivalently

$$\frac{\partial E_*(X, -\alpha(X - x) + y)}{\partial X}(x) = \frac{\partial E_*}{\partial X}(x, y) - \alpha \frac{\partial E_*}{\partial Y}(x, y) = 0.$$

This is exactly the condition on α above.

□

This lemma finishes the proof of Theorem 2.40: For a point P there is exactly one line with order at least 2 at P, namely the tangent line in P. □

2.10 THE PICARD GROUP

This section is devoted to finding "simple" representatives for each element of $\mathrm{Pic}(E)$. We show that $\mathrm{Pic}^0(E)$ is in bijective correspondence to the points of E, which motivates the definition of the group law on E via the group law of $\mathrm{Pic}^0(E)$.

Theorem 2.43 (Pushing poles and zeros to infinity)
Let $\Delta \in \mathrm{Div}(E)$. Then there is a unique point $P \in E$ such that

$$\Delta \sim \langle P\rangle + (\deg\Delta - 1)\langle\mathcal{O}\rangle.$$

Proof: The first step of the proof uses induction on the *norm* of the divisor. Define the norm of $\Delta = \sum_{P\in E} m_P\langle P\rangle$ by $|\Delta| = \sum_{P\in E\setminus\{\mathcal{O}\}} |m_P|$. We show that if $|\Delta| > 1$ then Δ can be replaced by an element of the same divisor class with smaller norm by adding the divisor of a suitable line. The details of the process are as follows:

- If there are two points P and Q with $m_P, m_Q > 0$ then subtract the divisor of the line l through P and Q. This reduces $|m_P|$ and $|m_Q|$ by 1. If $\mathrm{div}(l^*) = \langle P\rangle + \langle Q\rangle - 2\langle\mathcal{O}\rangle$, this is all. Otherwise $\mathrm{div}(l^*) = \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle\mathcal{O}\rangle$ and $|m_R|$ is at most increased by 1. In both cases $|\Delta|$ is reduced by at least 1.

- If there are two points P and Q with $m_P, m_Q < 0$, then add the divisor of the line l through P and Q.

In the remaining cases Δ is of the form $m\langle P\rangle - n\langle Q\rangle + o\langle\mathcal{O}\rangle$ with $m, n \geq 0$.

- If $m \geq 2$, subtract the divisor of the tangent line in P, which is $2\langle P\rangle + \langle R\rangle - 3\langle\mathcal{O}\rangle$. If $n \geq 2$, add the divisor of the tangent line in Q. In both cases $|\Delta|$ is decreased by at least 1.

- If $\Delta = \langle P\rangle - \langle Q\rangle + o\langle\mathcal{O}\rangle$, add the divisor of the line through Q and $\overline{Q}$, which is $\langle Q\rangle + \langle\overline{Q}\rangle - 2\langle\mathcal{O}\rangle$. The resulting divisor $\langle P\rangle + \langle\overline{Q}\rangle + (o-2)\langle\mathcal{O}\rangle$ has the same norm and falls under one of the previous cases, so its norm is reduced in the next step.

- If one of m or n is zero, we are finished.

The result of the process is a divisor $\Delta' \sim \Delta$ with norm at most 1. If $|\Delta'| = 0$, then $\Delta' = \langle\mathcal{O}\rangle + (\deg\Delta - 1)\langle\mathcal{O}\rangle$ is of the desired form. If $\Delta' = -\langle P\rangle + (\deg\Delta + 1)\langle\mathcal{O}\rangle$, we can add the divisor of the line through P and $\overline{P}$ to obtain the divisor $\langle\overline{P}\rangle + (\deg\Delta - 1)\langle\mathcal{O}\rangle$. This proves the existence.

For the uniqueness suppose $\Delta \sim \langle P\rangle - o\langle\mathcal{O}\rangle \sim \langle Q\rangle - o\langle\mathcal{O}\rangle$. Then $\langle P\rangle - \langle Q\rangle$ is a principal divisor. Proceed similarly as in the proof of the existence and subtract $\langle P\rangle + \langle\overline{P}\rangle - 2\langle\mathcal{O}\rangle$, which belongs to the line through P and $\overline{P}$, from this divisor. Then add $\langle\overline{P}\rangle + \langle Q\rangle + \langle R\rangle - 3\langle\mathcal{O}\rangle$, the divisor of the line through $\overline{P}$ and Q. If $P \neq Q$, then $\overline{P}$ and Q are not conjugate and R is finite by Corollary 2.39.

It follows that $\langle R\rangle - \langle O\rangle$ is principal, say to the rational function r. Since r has no finite poles, it is a polynomial by Proposition 2.34. But then r has only one finite zero, contradicting Corollary 2.33. □

Corollary 2.44 *For $\Delta \in \mathrm{Pic}^0(E)$ there is a unique $P \in E$ with $\Delta = \langle P\rangle - \langle O\rangle$. The maps*

$$\sigma : \mathrm{Pic}^0(E) \to E, \quad \Delta \mapsto P$$

and

$$\kappa : E \to \mathrm{Pic}^0(E), \quad P \mapsto \Delta$$

define inverse bijections.

2.11 THE GROUP LAW

We are finally ready to define the group law on elliptic curves. Our first approach is geometric and gives the rough idea, see Figure 2.2. To add two points P and Q, draw the line through the points (or the tangent line in P if $P = Q$). Determine the third point of intersection R of the line with the curve. Then $P + Q = \overline{R}$. There is also a geometric interpretation for $\overline{R}$: Draw the vertical line through R. Then the second point of intersection is $\overline{R}$ (the third point being O).

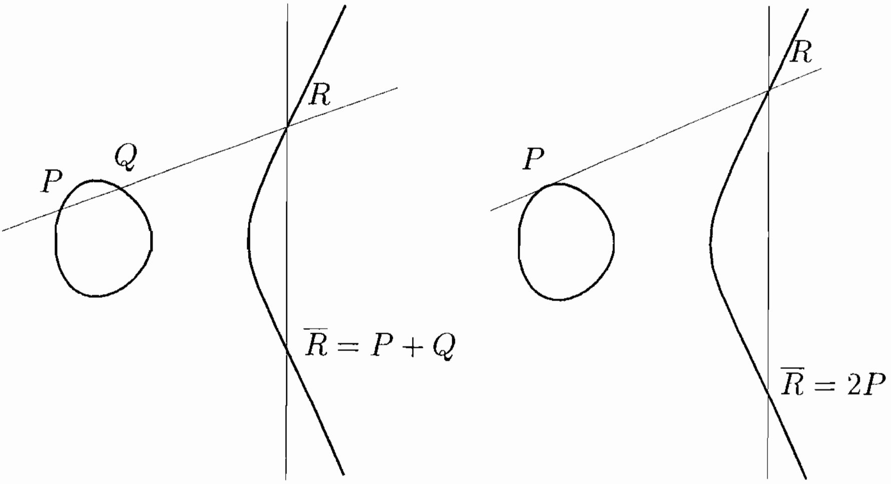

Figure 2.2. Adding two distinct points and doubling a point on an elliptic curve

We have two tasks now: First, verify that the so defined composition law turns E into a group. Second, make the definition operational by giving explicit algebraic formulae for the sum of two points. For both purposes it is necessary to formalise the definition of the composition law.

Definition 2.45 *On an elliptic curve a composition law* $+$ *can be defined as follows:*

- $\mathcal{O} + \mathcal{O} = \mathcal{O}$
- *For points* P *and* Q*, not both infinite, there is a unique line* l *and a unique point* R *with* $\operatorname{div} l^* = \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle \mathcal{O}\rangle$ *(Theorem 2.40). Let* $P + Q := \overline{R}$.

Theorem 2.46 $(E, +)$ *is an abelian group.*

Proof: Most of the properties of a group can be verified directly using the definition.

- $+$ is trivially commutative.
- $\mathcal{O}$ is the neutral element. If P is infinite, then $P + \mathcal{O} = \mathcal{O} = P$ by definition. Otherwise $P = (x, y, 1)$ and the line through P and $\mathcal{O}$ is $l = X - x$ with $\operatorname{div} l^* = \langle P\rangle + \langle \mathcal{O}\rangle + \langle \overline{P}\rangle - 3\langle \mathcal{O}\rangle$. So $P + \mathcal{O} = \overline{\overline{P}} = P$.
- $-P = \overline{P}$ is the inverse of P. For infinite P there is nothing to show. For $P = (x, y, 1)$, $\operatorname{div}((X - x)^*) = \langle P\rangle + \langle \overline{P}\rangle + \langle \mathcal{O}\rangle - 3\langle \mathcal{O}\rangle$, which shows that $P + \overline{P} = \overline{\mathcal{O}} = \mathcal{O}$.

Only the associativity causes difficulties. It can be proved geometrically, using the theory of algebraic curves (see [Fulton, 1969], p. 125). We are not following this approach, but instead prove that $(E, +)$ is isomorphic to $\operatorname{Pic}^0(E)$. Recall the bijection $\kappa : E \to \operatorname{Pic}^0(E)$, $P \mapsto \langle P\rangle - \langle \mathcal{O}\rangle$ from Section 2.10. All what is left to show is that κ is compatible with the composition law on E, i.e.

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

This is trivial for $P = Q = \mathcal{O}$, so we can suppose at least one of P and Q to be finite. Let l be the line with $\operatorname{div} l^* = \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle \mathcal{O}\rangle$ such that $P + Q = \overline{R}$. We have to show that

$$\langle \overline{R}\rangle - \langle \mathcal{O}\rangle \sim \langle P\rangle + \langle Q\rangle - 2\langle \mathcal{O}\rangle,$$

or that $\langle P\rangle + \langle Q\rangle - \langle \overline{R}\rangle - \langle \mathcal{O}\rangle$ is a principal divisor. If l' is the line with divisor $\langle R\rangle + \langle \overline{R}\rangle - 2\langle \mathcal{O}\rangle$, then $\operatorname{div} \frac{l^*}{l'^*} = \langle P\rangle + \langle Q\rangle - \langle \overline{R}\rangle - \langle \mathcal{O}\rangle$ is principal. □

Corollary 2.47

1. $(E, +)$ *is isomorphic to* $\operatorname{Pic}^0(E)$.
2. *If* E *and* E' *are isomorphic, then so are* $(E, +)$ *and* $(E', +)$.

Proof: The first assertion has just been proved. The second one follows from the fact that an isomorphism between elliptic curves is compatible with their divisor theories, see the discussion in Section 2.8. □

For an implementation it is now important to compute the coordinates of $P+Q$ from the coordinates of P and Q. Let $P=(x_1,y_1,1)$ and $Q=(x_2,y_2,1)$ (if one of P and Q is infinite, nothing is to do). Check first if $Q=\overline{P}$, i.e. $x_1=x_2$ and $y_2=-y_1-a_1x_1-a_3$. Then $P+Q=\mathcal{O}$. Otherwise the line l through P and Q has the equation

$$l = Y - (\lambda X + \mu)$$

with

$$\begin{aligned} \lambda &= \begin{cases} \dfrac{y_2-y_1}{x_2-x_1} & \text{if } P \neq Q \\[2ex] \dfrac{3x_1^2+2a_2x_1+a_4-a_1y_1}{2y_1+a_1x_1+a_3} & \text{if } P = Q\text{, see the example} \\ & \text{after Definition 2.41} \end{cases} \\ \mu &= y_1 - \lambda x_1 \end{aligned}$$

The third point of intersection $R=(x_3,y_3,1)$ of l with E can be computed via the relation

$$-(X-x_1)(X-x_2)(X-x_3) = E_*(X, \lambda X + \mu)$$

proved in Lemma 2.38. The coefficients of X^2 on both sides of this equation are $x_1+x_2+x_3$ and $\lambda^2+a_1\lambda-a_2$, respectively, so $x_3=\lambda^2+a_1\lambda-a_2-x_1-x_2$ and $y_3=\lambda x_3+\mu$. Finally $P+Q=\overline{R}=(x_3,-y_3-a_1x_3-a_3)=(x_3,-(\lambda+a_1)x_3-a_3-\mu)$.

Tables 2.3 to 2.5 summarise the addition formulae for the general case and the normal forms in characteristic 2. They record the sum (x_3,y_3) of $P=(x_1,y_1)=(x,y)$ and $Q=(x_2,y_2)$ for $P\neq -Q$. The only formula which deserves explanation is that for the X-coordinate of twice a point in Table 2.4. Specialising the formula of Table 2.3 yields

$$\begin{aligned} x_3 &= \left(\frac{x^2+y}{x}\right)^2 + \frac{x^2+y}{x} + a_2 \\ &= \frac{x^4+y^2+x^3+xy+a_2x^2}{x^2} \\ &= \frac{x^4+a_6}{x^2} \end{aligned}$$

since (x,y) lies on E and hence $y^2+xy+x^3+a_2x^2=a_6$.

$$x_3 = \begin{cases} \left(\dfrac{y_2 - y_1}{x_2 - x_1}\right)^2 + a_1\left(\dfrac{y_2 - y_1}{x_2 - x_1}\right) - a_2 - x_1 - x_2 & \text{if } P \neq Q \\[2ex] \left(\dfrac{3x^2 + 2a_2x + a_4 - a_1y}{2y + a_1x + a_3}\right)^2 + a_1\left(\dfrac{3x^2 + 2a_2x + a_4 - a_1y}{2y + a_1x + a_3}\right) \\ \qquad -a_2 - 2x & \text{if } P = Q \end{cases}$$

$$y_3 = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1}(x_1 - x_3) - y_1 - (a_1x_3 + a_3) & \text{if } P \neq Q \\[2ex] \dfrac{3x^2 + 2a_2x + a_4 - a_1y}{2y + a_1x + a_3}(x - x_3) - y - (a_1x_3 + a_3) & \text{if } P = Q \end{cases}$$

Table 2.3. Addition formulae for $Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6$

$$x_3 = \begin{cases} \left(\dfrac{y_1 + y_2}{x_1 + x_2}\right)^2 + \left(\dfrac{y_1 + y_2}{x_1 + x_2}\right) + a_2 + x_1 + x_2 & \text{if } P \neq Q \\[2ex] x^2 + \dfrac{a_6}{x^2} & \text{if } P = Q \end{cases}$$

$$y_3 = \begin{cases} \dfrac{y_1 + y_2}{x_1 + x_2}(x_1 + x_3) + y_1 + x_3 & \text{if } P \neq Q \\[2ex] \dfrac{x^2 + y}{x}x_3 + x^2 + x_3 & \text{if } P = Q \end{cases}$$

Table 2.4. Addition formulae for char $K = 2$, $j \neq 0$, $Y^2 + XY = X^3 + a_2X^2 + a_6$

$$x_3 = \begin{cases} \left(\dfrac{y_1 + y_2}{x_1 + x_2}\right)^2 + x_1 + x_2 & \text{if } P \neq Q \\[2ex] \left(\dfrac{x^2 + a_4}{a_3}\right)^2 & \text{if } P = Q \end{cases}$$

$$y_3 = \begin{cases} \dfrac{y_1 + y_2}{x_1 + x_2}(x_1 + x_3) + y_1 + a_3 & \text{if } P \neq Q \\[2ex] \dfrac{x^2 + a_4}{a_3}(x + x_3) + y + a_3 & \text{if } P = Q \end{cases}$$

Table 2.5. Addition formulae for char $K = 2$, $j = 0$, $Y^2 + a_3 Y = X^3 + a_4 X + a_6$

Up to now we restricted our attention to elliptic curves over algebraically closed fields. However, the interesting fields, which allow effective computations (and thus can be used for cryptographical purposes), are the finite fields. The algebraic formulae for the coordinates of $P + Q$ and $-P$ show that if we add or invert points with coordinates in a subfield $k \subseteq K$ we again obtain a point with coordinates in k. So we get a first result on arbitrary fields:

Corollary 2.48 *Let K be a field, not necessarily algebraically closed, and E an elliptic curve defined over K. Then $(E, +)$ is a group, where the composition $+$ is defined by the above algebraic formulae.*

Example. We choose as underlying field the field $\mathbb{F}_4$ with four elements. $\mathbb{F}_4 = \{0, 1, \alpha, \alpha + 1\}$ where α generates the cyclic multiplicative group of $\mathbb{F}_4$, so $\alpha^2 = \alpha^{-1} = \alpha + 1$, $(\alpha + 1)^2 = \alpha^4 = \alpha$ and $\alpha(\alpha + 1) = \alpha^3 = 1$. As curve we choose $E : Y^2 + Y = X^3 + X + 1$.

Since $j = 0$, there are no points of order 2 (see Section 2.5). Which finite points $P = (x, y)$ lie on E? If $x = 0$ or $x = 1$, then $y^2 + y = 1$, so $y = \alpha$ or $y = \alpha + 1$. If $x = \alpha$ resp. $x = \alpha + 1$, then $y^2 + y = x$, and there is no solution.

So E consists of the points $P_1 = (0, \alpha)$, $P_2 = (0, \alpha + 1)$, $P_3 = (1, \alpha)$, $P_4 = (1, \alpha + 1)$ and $\mathcal{O}$, and the group must be cyclic of order 5 and be generated by any of the P_i. We verify this directly by computing the multiples of P_1. Recall from Table 2.5 the relevant formulae for this special curve with $P = (x_1, y_1)$, $Q = (x_2, y_2) \neq \overline{P}$, $R = P + Q = (x_3, y_3)$:

$$\begin{aligned}
\overline{P} &= (x_1, y_1 + 1) \\
x_3 &= \begin{cases} \left(\dfrac{y_1 + y_2}{x_1 + x_2}\right)^2 + x_1 + x_2 & \text{for } P \neq Q \\ x_1^4 + 1 = x_1 + 1 & \text{for } P = Q \end{cases} \\
y_3 &= \begin{cases} \left(\dfrac{y_1 + y_2}{x_1 + x_2}\right)^3 + \left(\dfrac{y_1 + y_2}{x_1 + x_2}\right) x_2 + y_1 + 1 & \text{for } P \neq Q \\ x_1^2 + y_1 & \text{for } P = Q \end{cases}
\end{aligned}$$

So the multiples of P_1 are computed as follows:

$$\begin{aligned}
0P_1 &= \mathcal{O} \\
-P_1 &= \overline{P_1} = (0, \alpha + 1) = P_2 \\
2P_1 &= (1, \alpha) = P_3 \\
-2P_1 &= \overline{2P_1} = (1, \alpha + 1) = P_4 \\
3P_1 &= P_1 + P_3 = P_4 = -2P_1 \text{ as expected.}
\end{aligned}$$

3 ELLIPTIC CURVES OVER FINITE FIELDS

Et l'Infini terrible effara ton œil bleu!

—Rimbaud

We have verified in the previous chapter that the points on an elliptic curve over an arbitrary field form a group, which can be used to implement the public key cryptosystems presented in the first chapter. Since by the algebraic formulae the group operations eventually amount to computations in the field where the elliptic curve is defined, one has to choose a field with an efficiently implementable arithmetic. Basically, this requirement narrows down to the finite fields. (While the rational numbers and more generally number fields also allow exact computations, they have two drawbacks: First, numbers may become arbitrarily big, which destroys the efficiency of the operations. And more important, the discrete logarithm problem on elliptic curves over these fields is easy to solve.) So during this chapter, we consider the following situation:

Let $k = \mathbb{F}_q$ be the finite field with q elements and prime characteristic p, and $K = \overline{k}$ be its algebraic closure. Let E be an elliptic curve which is defined over k, i.e. whose defining coefficients a_1, a_3, a_2, a_4 and a_6 lie in k. As before we denote by E the group of points on the curve with coordinates in K. The group of k-rational points, i.e. the group of points on E with coordinates in k, in which we are eventually interested, is denoted by E_k. Since k is finite, there are only finitely many possibilities for the X- and Y-coordinates of points, and

E_k is a finite abelian group. We will see in Chapter 4 that it is mainly the exact cardinality of E_k which determines the security of a cryptosystem built in this group. The biggest part of this chapter is devoted to the proof of Hasse's famous theorem, which gives an estimate on the cardinality of E_k, stating that the elliptic curve group has roughly as many elements as k itself. Precisely,

$$|q+1-|E_{\mathbb{F}_q}|| \leq 2\sqrt{q}.$$

We follow very closely the excellent report [Charlap and Robbins, 1988], occasionally putting a different emphasis. However, we take care to prove all theorems for characteristic 2 as well.

While the results of the first seven sections hold in full generality for any finite or infinite field k, we will apply them to finite fields only: namely to prove Hasse's Theorem in Section 3.8 and to compute the exact cardinality of E_k in Chapter 5. In the last sections of this chapter we present some results on special classes of elliptic curves over finite fields and on the group structure of E_k.

3.1 RATIONAL MAPS AND ENDOMORPHISMS

With rational functions we have up to now considered mappings from an elliptic curve E to the underlying fields k resp. K. We now define maps from E to itself. In affine notation we examine pairs of rational functions $\alpha = (\alpha_1, \alpha_2) \in K(E) \times K(E)$ such that for a point $P \in E$ its image $\alpha(P) := (\alpha_1(P), \alpha_2(P))$ lies on E again. Formally we need

$$(E \circ \alpha)(P) = E(\alpha(P)) = 0 \quad \forall P \in E.$$

This amounts to requiring that $E(\alpha) = E \circ \alpha$ is the zero rational function, and so we can formulate the following definition:

Definition 3.1 *A* rational map *is a point on the elliptic curve* $E_{K(E)}$. *To avoid confusion we denote the infinite point on this curve by* [0].

This definition allows a straightforward generalisation to rational maps between distinct elliptic curves E and E' (cf. Definition 5.12). The reader is invited to check the following results in the more general context.

For rational maps we prefer the affine notation to the projective one, but then we have to define the image of a point $P \in E$ under a rational map α. First fix the image of any point under the zero map [0] as the zero element $\mathcal{O}$ of E. Otherwise $\alpha = (\alpha_1, \alpha_2)$. If α_1 and α_2 are regular at P, let $\alpha(P) = (\alpha_1(P), \alpha_2(P))$; if α_1 and α_2 are not regular at P, let $\alpha(P) = \mathcal{O}$. Note that α_1 and α_2 are either both regular or both not regular at P since they satisfy the equality

$$\alpha_2^2 + a_1\alpha_1\alpha_2 + a_3\alpha_2 = \alpha_1^3 + a_2\alpha_1^2 + a_4\alpha_1 + a_6.$$

Assume that $\operatorname{ord}_P \alpha_1 < 0$, $\operatorname{ord}_P \alpha_2 \geq 0$; then with Proposition 2.14 the left hand side of the equation has at least the order $\operatorname{ord}_P \alpha_1$ in P, the right hand side has order $3\,\mathrm{ord}_P\,\alpha_1 < \mathrm{ord}_P\,\alpha_1$, a contradiction. In the converse case $\mathrm{ord}_P\,\alpha_2 < 0$, $\mathrm{ord}_P\,\alpha_1 \geq 0$, a similar argument holds.

Since rational maps are defined as points on a special elliptic curve they form a group under the usual addition law. It is now a natural question to ask what the image of a point under the sum of two rational maps is. The answer is given by the following theorem, whose suggestive notation hides that we need a long breath for proving it.

Theorem 3.2 *Let α and β be rational maps. Then*

$$(\alpha+\beta)(P) = \alpha(P) + \beta(P) \quad \forall P \in E.$$

Proof: We have to distinguish several cases, following the definition of the addition on an elliptic curve. Notice that the additions on either side of the equation take place on curves over different fields, namely on $E_{K(E)}$ and E_K, and that even if $\alpha \neq \pm\beta$, it is possible for a specific point P that $\alpha(P) = \pm\beta(P)$, which complicates the case distinction. Let P be a fixed point on E.

- If one of α and β is $[0]$, then the assertion is clear.

Otherwise we can write $\alpha = (\alpha_1, \alpha_2)$ and $\beta = (\beta_1, \beta_2)$.

- If $\alpha = -\beta$, then $\alpha_1 = \beta_1$ and $\alpha_2 = -\beta_2 - a_1\beta_1 - a_3$. It follows that either $\alpha_1 = \beta_1$ is not regular at P and $\alpha(P) + \beta(P) = \mathcal{O} + \mathcal{O} = \mathcal{O} = [0](P) = (\alpha+\beta)(P)$, or $\alpha_1(P) = \beta_1(P)$ and $\alpha_2(P) = -\beta_2(P) - a_1\beta_1(P) - a_3$, i.e. $\alpha(P) = -\beta(P)$ and $\alpha(P) + \beta(P) = \mathcal{O}$ as desired.

In the remaining cases, $\alpha + \beta \neq [0]$, so we write $\alpha + \beta = \gamma = (\gamma_1, \gamma_2)$.

- Assume that $\alpha \neq \beta$. Then

$$\gamma_1 = -\alpha_1 - \beta_1 + \lambda^2 + \lambda a_1 - a_2 \tag{3.1}$$

and

$$\gamma_2 = \lambda(\alpha_1 - \gamma_1) - \beta_1 - (a_1\gamma_1 + a_3) \tag{3.2}$$

with

$$\lambda = \frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1}. \tag{3.3}$$

Several cases have to be distinguished:

  - $\alpha(P)$ and $\beta(P)$ are finite and distinct, furthermore $\alpha(P) \neq -\beta(P)$. Then substituting P into (3.1) to (3.3) yields the usual addition formula for $\alpha(P)$ and $\beta(P)$ and shows that $\gamma(P) = \alpha(P) + \beta(P)$.
  - $\alpha(P)$ and $\beta(P)$ are finite and distinct, but $\alpha(P) = -\beta(P)$. Then $\alpha_1(P) = \beta_1(P)$ and $\alpha_2(P) \neq \beta_2(P)$, so λ has a pole at P. By Proposition 2.14, γ_1 has a pole at P, too, and $\gamma(P) = \mathcal{O} = \alpha(P) + \beta(P)$.

- $\alpha(P) = \beta(P)$ is finite and not of order 2, i.e. $\alpha_1(P) = \beta_1(P) =: x$, $\alpha_2(P) = \beta_2(P) =: y$ and $2y + a_1x + a_3 \neq 0$. Then the formulae for adding α and β resp. $\alpha(P)$ and $\beta(P)$ only differ by the definition of λ. In the latter case we compute with

$$\lambda' = \frac{3\alpha_1^2 + 2a_2\alpha_1 + a_4 - a_1\alpha_2}{2\alpha_2 + a_1\alpha_1 + a_3},$$

evaluated at P. So we have to show that $\lambda(P) = \lambda'(P)$. We compute

$$\begin{aligned}
\lambda &= \frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1} \\
&= \frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1}\,\frac{\beta_2 + \alpha_2 + a_1\alpha_1 + a_3}{\beta_2 + \alpha_2 + a_1\alpha_1 + a_3} \\
&= \frac{(\beta_2^2 + a_1\beta_1\beta_2 + a_3\beta_2) - a_1\beta_2(\beta_1 - \alpha_1) - (\alpha_2^2 + a_1\alpha_1\alpha_2 + a_3\alpha_2)}{(\beta_1 - \alpha_1)(\beta_2 + \alpha_2 + a_1\alpha_1 + a_3)} \\
&= \frac{(\beta_1^3 - \alpha_1^3) + a_2(\beta_1^2 - \alpha_1^2) + (a_4 - a_1\beta_2)(\beta_1 - \alpha_1)}{(\beta_1 - \alpha_1)(\beta_2 + \alpha_2 + a_1\alpha_1 + a_3)} \\
&\quad \text{because } \beta \text{ and } \alpha \text{ satisfy } E \text{ by the definition of a rational map} \\
&= \frac{(\beta_1^2 + \beta_1\alpha_1 + \alpha_1^2) + a_2(\beta_1 + \alpha_1) + (a_4 - a_1\beta_2)}{\beta_2 + \alpha_2 + a_1\alpha_1 + a_3}
\end{aligned}$$

Then with $\alpha(P) = \beta(P) = (x, y)$ we have that

$$\begin{aligned}
\lambda(P) &= \frac{3x^2 + 2a_2x + a_4 - a_1y}{2y + a_1x + a_3} \\
&= \lambda'(P).
\end{aligned}$$

- $\alpha(P) = \beta(P)$ is finite and of order 2. As in the previous case we can write

$$\lambda = \frac{(\beta_1^2 + \beta_1\alpha_1 + \alpha_1^2) + a_2(\beta_1 + \alpha_1) + (a_4 - a_1\beta_2)}{\beta_2 + \alpha_2 + a_1\alpha_1 + a_3}.$$

Now the denominator has the value $\frac{\partial E}{\partial Y}(\alpha(P))$ and the numerator has the value $\frac{\partial E}{\partial X}(\alpha(P))$ in P. Since $\alpha(P)$ is of order two, $\frac{\partial E}{\partial Y}(\alpha(P)) = 0$, and so $\frac{\partial E}{\partial X}(\alpha(P)) \neq 0$ because E is non-singular. But then λ has a pole in P, and so does γ by Proposition 2.14. So $\gamma(P) = \mathcal{O} = 2\alpha(P)$.

- Exactly one of $\alpha(P)$ and $\beta(P)$ is infinite, say $\alpha(P) = \mathcal{O}$. Then $\alpha_1 = u^{d_1}\alpha_1'$ and $\alpha_2 = u^{d_2}\alpha_2'$ where u is a uniformising parameter for P, $d_1, d_2 < 0$ and α_1' and α_2' are regular and not zero in P.

Our first goal is to show that $\gamma(P)$ is finite. From $E(\alpha_1, \alpha_2) = 0$ and Proposition 2.14 we deduce that

$$\begin{aligned} \min\{2d_2, d_1 + d_2\} &\leq \operatorname{ord}_P(\alpha_2^2 + a_1\alpha_1\alpha_2 + a_3\alpha_2) \\ &= \operatorname{ord}_P(\alpha_1^3 + a_2\alpha_1^2 + a_4\alpha_1 + a_6) \\ &= 3d_1, \end{aligned}$$

which implies $d_2 < d_1$ by the negativity of d_1 and d_2. Again by Proposition 2.14 it follows that $2d_2 = 3d_1$. By substituting formula (3.3) into (3.1) and repeatedly replacing occurrences of α_2^2 by $\alpha_1^3 + a_2\alpha_1^2 + a_4\alpha_1 + a_6 - a_1\alpha_1\alpha_2 - a_3\alpha_2$ and β_2^2 by the corresponding term we obtain the following new formula for γ_1:

$$\begin{aligned} \gamma_1 &= \frac{(\alpha_1\beta_1 + a_4)(\alpha_1 + \beta_1) - a_1(\alpha_1\beta_2 + \beta_1\alpha_2) + 2a_2\alpha_1\beta_1}{(\alpha_1 - \beta_1)^2} \\ &\quad - \frac{a_3(\alpha_2 + \beta_2) + 2\alpha_2\beta_2 - 2a_6}{(\alpha_1 - \beta_1)^2} \end{aligned}$$

As $\operatorname{ord}_P(\beta_1)$, $\operatorname{ord}_P(\beta_2) \geq 0$ it follows by Proposition 2.14 that

$$\begin{aligned} \operatorname{ord}_P(\gamma_1) &\geq \min\{2d_1, d_2\} - 2d_1 = \min\{0, d_2 - 2d_1\} \\ &= \min\left\{0, -\frac{1}{2}d_1\right\} = 0. \end{aligned}$$

Taking into account that the group law for rational maps is associative and commutative we write $\alpha = \gamma - \beta$ where both γ and β are finite at P. Applying the cases proven above we find that $\alpha(P) = \gamma(P) - \beta(P)$.

- $\alpha(P) = \beta(P) = \mathcal{O}$. Assume $\gamma(P) \neq \mathcal{O}$ and write $\alpha = \gamma - \beta$. By the previous case, $\mathcal{O} = \alpha(P) = \gamma(P) - \beta(P) = \gamma(P)$, a contradiction.

- Finally assume that $\alpha = \beta \neq -\beta$. Then γ_1 satisfies (3.1) with

$$\lambda = \frac{3\alpha_1^2 + 2a_2\alpha_1 + a_4 - a_1\alpha_2}{2\alpha_2 + a_1\alpha_1 + a_3}.$$

There are two cases to consider:

- $\alpha(P) \neq \mathcal{O}$
  Then $\gamma(P) = 2\alpha(P)$ follows by substituting P into (3.1), which yields exactly the duplication formula for points on E.

- $\alpha(P) = \mathcal{O}$
  Assume $\gamma(P) \neq \mathcal{O}$. Then by writing $\alpha = \gamma - \alpha$ it follows by one of the previous cases that $\mathcal{O} = \gamma(P)$, a contradiction.

□

Example. First trivial examples of rational maps are provided by the identity map $\mathrm{id} = (X, Y)$ and the constant maps $c_Q = (x, y)$ for $Q = (x, y) \in E$. These maps can be used to define a more important example, namely the translation by Q:

$$\tau_Q : P \mapsto P + Q.$$

Theorem 3.2 shows that the pointwise definition of τ_Q translates into the equality

$$\tau_Q = \mathrm{id} + c_Q$$

where the right hand side is a rational map. So the translation maps are rational.

Proposition 3.3 *A rational map is either constant or surjective.*

Proof: We first show the analogous assertion for rational functions. Let r be a non-constant rational function. Then r must have a zero, and the same argument applied to $r - x$ for $x \in K$ shows that r takes all elements of K as values.

Now consider a rational map $\alpha = (\alpha_1, \alpha_2)$. If α_1 is constant, then α_2 can take only finitely many values, namely the roots of $E(\alpha_1, Y) \in K[Y]$ in K. So α_2 is not surjective and hence constant, and α is constant, too. Otherwise α_1 is surjective; it has a zero and consequently a pole by Theorem 2.28, so α takes the value $\mathcal{O}$. Apply the same reasoning to $\tau_{-Q} \circ \alpha$ to see that Q is a value of α for all $Q \in E$. □

Definition 3.4 *An* endomorphism *or* isogeny *is a rational map which is furthermore a group homomorphism. The set of endomorphisms of E is denoted by* $\mathrm{End}(E)$.

Proposition 3.5 $\mathrm{End}(E)$ *is a ring in which the multiplication is given by composition.*

Proof: The only point to verify is the distributive law. So let α, β and γ be endomorphisms.

$$\begin{aligned}(\alpha \circ (\beta + \gamma))(P) &= \alpha((\beta + \gamma)(P)) \\ &= \alpha(\beta(P) + \gamma(P)) \quad \text{by Theorem 3.2} \\ &= \alpha(\beta(P)) + \alpha(\gamma(P)) \quad \text{because } \alpha \text{ is a homomorphism} \\ &= (\alpha \circ \beta)(P) + (\alpha \circ \gamma)(P).\end{aligned}$$

It follows that $\alpha \circ (\beta + \gamma) = \alpha \circ \beta + \alpha \circ \gamma$. That $(\alpha + \beta) \circ \gamma = \alpha \circ \gamma + \beta \circ \gamma$, is even easier to verify and holds for any rational maps. □

Example. Of the constant and translation maps considered so far only $c_{\mathcal{O}} = [0]$ and $\tau_{\mathcal{O}} = \mathrm{id}$ are endomorphisms. More interesting examples are provided by the multiplication-by-m-maps

$$[m] : P \mapsto mP$$

for $m \in \mathbb{Z}$. Since the pointwise relations $mP = (m-1)P + P$ for $m > 0$ and $mP = -((-m)P)$ for $m < 0$ translate by Theorem 3.2 into the equalities $[m] = [m-1] + \mathrm{id}$ for $m > 0$ and $[m] = -[-m]$ for $m < 0$, we see recursively that $[m]$ is rational. Furthermore it is an endomorphism since $m(P+Q) = mP + mQ$.

Corollary 3.6 End(E) *is a* $\mathbb{Z}$*-algebra, where the multiplication by* $m \in \mathbb{Z}$ *is given by the composition with* $[m]$.

Proof: We have already shown in Proposition 3.5 that End(E) is a ring. Then the example of the multiplication by m shows that $\mathbb{Z}$ is a subring, so End(E) carries a natural structure of $\mathbb{Z}$-module. □

Definition 3.7 *If* End(E) *contains another endomorphism than the multiplications by* m*, then* E *is said to have* complex multiplication.

If $k = \mathbb{F}_q$ is a finite field, then E has complex multiplication, namely by the *Frobenius endomorphism* $\varphi = (X^q, Y^q)$: If $P = (x,y)$ is a point on E, then $E(\varphi(P)) = E(x^q, y^q) = E(x,y)^q = 0$, so $\varphi(P) \in E$. Note that in the step $E(x^q, y^q) = E(x,y)^q$ we need that E is defined over k so that $a_i^q = a_i$ for the coefficients of E. The same argument shows that φ is an endomorphism, since it is compatible with the rational addition formulae. The Frobenius endomorphism is a key ingredient to the proof of Hasse's theorem. Another important ingredient is the structure of End(E) as a $\mathbb{Z}$-module when the endomorphisms are restricted to m-torsion points, which in turn are closely related to the multiplication by m.

Definition 3.8 *Let* m *be an integer and* P *a point on* E*. Then* P *is called an* m-torsion point *if* $mP = \mathcal{O}$*. The set of all* m*-torsion points is denoted by* $E[m]$*. The point* P *is* of order m *if* $mP = \mathcal{O}$ *and* $m'P \neq \mathcal{O}$ *for* $0 < m' < m$.

Notice that the m-torsion points form exactly the kernel of $[m]$.

Theorem 3.9 *Let* m *be a non-zero integer. Then* $E[m]$ *is finite and* $[m] \neq [0]$*, so we can write* $[m] = (g_m, h_m)$*. The rational functions* g_m *and* h_m *have poles exactly at the points of* $E[m]$.

Proof: First note that $[m] \neq [0]$ is equivalent to $|E[m]| < \infty$: If $[m] = [0]$, then $E[m] = \ker[m] = E$ is infinite. If $[m] = (g_m, h_m) \neq [0]$, then $[m](P) = mP = \mathcal{O}$ means that g_m and h_m have a pole at P. Since a rational function can have only a finite number of poles, $E[m]$ must be finite. So we are left with showing that $[m] \neq [0]$. Obviously it is sufficient to consider only positive values of m because $E[-m] = E[m]$. We distinguish several cases:

1. The theorem is clear for $m = 1$.

2. For $m = 2$ we have to show that $[2] \neq [0]$ or equivalently that $[1] \neq [-1]$, which again is clear.

3. Suppose now that m is an odd prime and $p \neq 2$. Recall from Section 2.5 that there is a point P of order 2. Then

$$mP = (m-1)P + P = \mathcal{O} + P = P \neq \mathcal{O}$$

   and $[m] \neq [0]$.

4. Let m be an odd prime and $p = 2$. If $j \neq 0$, then there is a point of order 2 by Section 2.5, and the proof of the previous case carries over immediately. Otherwise we show that there is a point of order 3 and apply a similar argument. By Section 2.3 we can assume that E is of the form $Y^2 + a_3Y = X^3 + a_4X + a_6$. Recall the duplication formula for $P = (x, y)$:

$$X(2P) = \frac{x^4 + a_4^2}{a_3^2}$$

   and the formula for the opposite

$$X(-P) = X(P).$$

   Now the 3-torsion points are exactly the points P with $2P = \pm P$, i.e. $X(2P) = X(P)$. Choose $x \in K$ such that $\frac{x^4+a_4^2}{a_3^2} = x$ and solve $E(x, Y) = 0$ with $y \in K$. Then $P = (x, y)$ is a point of order 3. If $m \neq 3$, then $mP \neq \mathcal{O}$ and $[m] \neq [0]$ as in the previous case. Finally note that

$$E[3] = \{(x, y) : x^4 + a_3^2x + a_4^2 = 0,\ E(x, y) = 0\}$$

   is finite: There are at most four possibilities for x, and to each of these values correspond at most two values for y. So $[3] \neq [0]$.

5. If m is composite we proceed by induction on the number of its prime factors. Let d be a proper divisor of m, and consider the group homomorphism

$$\rho : E[m] \to E[d], \quad P \mapsto \frac{m}{d}P.$$

   Its image is finite as a subgroup of $E[d]$ by the induction hypothesis, and so is its kernel $E[\frac{m}{d}]$. It follows that $|E[m]| = |\operatorname{im}\rho||\ker\rho|$ is finite, too, so $[m] \neq [0]$.

□

Corollary 3.10 *If $m \neq n$, then $[m] \neq [n]$.*

Proof: $[m] = [n]$ implies $[m - n] = [0]$, so $m - n = 0$ by the theorem above. □

We trivially have $g_1 = X$ and $h_1 = Y$, and the corollary allows us to compute g_m and h_m explicitly by induction. To obtain g_2 and h_2 we apply the duplication formula for $[2] = [1] + [1]$:

$$\begin{aligned} g_2 &= -2X + \lambda^2 + a_1\lambda - a_2 \\ h_2 &= -\lambda(g_2 - X) - a_1 g_2 - a_3 - Y \\ \lambda &= \frac{3X^2 + 2a_2X + a_4 - a_1Y}{2Y + a_1X + a_3} \end{aligned}$$

If $m > 2$, then $[m-1] \neq [1]$ by the corollary, and we use the generic addition formula to compute $[m] = [m-1] + [1]$:

$$\begin{aligned} g_m &= -g_{m-1} - X + \lambda^2 + a_1\lambda - a_2 \\ h_m &= -\lambda(g_m - X) - a_1 g_m - a_3 - Y \\ \lambda &= \frac{h_{m-1} - Y}{g_{m-1} - X} \end{aligned}$$

3.2 RAMIFICATION INDEX AND DEGREE

To a rational map α corresponds in a natural way a map

$$\alpha^* : K(E) \to K(E), \quad r \mapsto r \circ \alpha,$$

which we investigate in this section. A first observation is the following consequence of Proposition 3.3.

Proposition 3.11 *If α is a non-constant rational map, then α^* is an injective field homomorphism.*

Proof: It is trivial to see that α^* is a field homomorphism. The injectivity of α^* is a direct consequence of the surjectivity of α: If $r \circ \alpha = s \circ \alpha$, then $r(\alpha(P)) = s(\alpha(P))$ for all $P \in E$. As $\alpha(P)$ varies over all points of E, it follows that $r(Q) = s(Q)$ for all $Q \in E$, which means that $r = s$. □

Fixing a point P and taking u as a uniformising parameter in $\alpha(P)$, one might expect that $u \circ \alpha$ has order 1 in P since u has order 1 in $\alpha(P)$. But this is not true, as examples reveal. This observation leads to the following definition.

Definition 3.12 *Let α be a non-constant rational map, $P \in E$ and u a uniformising parameter at $\alpha(P)$. Then the number*

$$e_\alpha(P) := \mathrm{ord}_P(u \circ \alpha)$$

is called the ramification index *of α at P. If $e_\alpha > 1$, α is said to be* ramified *at P, otherwise it is* unramified *at P. α is called* unramified *if it is unramified at all points of E.*

Note that the ramification index is independent of the choice of u. Let u' be another uniformising parameter at $\alpha(P)$, so that $\frac{u'}{u}$ is regular and not zero in $\alpha(P)$ and $\frac{u'}{u} \circ \alpha$ is regular and not zero in P. Then

$$\begin{aligned} \mathrm{ord}_P(u' \circ \alpha) &= ord_P\left(\left(u\frac{u'}{u}\right) \circ \alpha\right) \\ &= ord_P(u \circ \alpha) + ord_P\left(\frac{u'}{u} \circ \alpha\right) \\ &= ord_P(u \circ \alpha) \end{aligned}$$

We would like to define another map α^* on divisors. This map should be "the same" as α^* on rational functions in the sense that it commutes with the formation of principal divisors. This property holds for the map

$$\alpha^* : \mathrm{Div}(E) \to \mathrm{Div}(E), \quad \langle Q \rangle \mapsto \sum_{P \in \alpha^{-1}(Q)} e_\alpha(P)\langle P \rangle,$$

$\mathbb{Z}$-linearly extended to the whole of $\mathrm{Div}(E)$.

Proposition 3.13 *The following diagram commutes for a non-constant rational map α:*

$$\begin{tikzcd} K(E) \arrow[r, "\alpha^*"] \arrow[d, "\mathrm{div}"] & K(E) \arrow[d, "\mathrm{div}"] & r \arrow[r] \arrow[d] & r \circ \alpha \arrow[d] \\ \mathrm{Div}(E) \arrow[r, "\alpha^*"] & \mathrm{Div}(E) & \mathrm{div}\, r \arrow[r] & \mathrm{div}(r \circ \alpha) \end{tikzcd}$$

Lemma 3.14 *Let α be a non-constant rational map, r a rational function and $P \in E$. Then*

$$\mathrm{ord}_P(r \circ \alpha) = e_\alpha(P)\ \mathrm{ord}_{\alpha(P)}(r).$$

Proof of the lemma: Let u be a uniformising parameter at $\alpha(P)$. Notice that for $r = u$ the lemma is exactly the definition of the ramification index. Write $r = u^d r_1$ with a rational function r_1 which is regular and not zero in $\alpha(P)$. Then

$$\begin{aligned} \mathrm{ord}_P(r \circ \alpha) &= d\,\mathrm{ord}_P(u \circ \alpha) + \mathrm{ord}_P(r_1 \circ \alpha) \\ &= d\, e_\alpha(P) \end{aligned}$$

□

Proof of the proposition:

$$\begin{aligned}
\operatorname{div}(r \circ \alpha) &= \sum_{P \in E} \operatorname{ord}_P(r \circ \alpha)\langle P\rangle \\
&= \sum_{P \in E} e_\alpha(P) \operatorname{ord}_{\alpha(P)}(r)\langle P\rangle \quad \text{by the lemma} \\
&= \sum_{Q \in E} \operatorname{ord}_Q(r) \sum_{P \in \alpha^{-1}(Q)} e_\alpha(P)\langle P\rangle \\
&= \sum_{Q \in E} \operatorname{ord}_Q(r)\, \alpha^*(\langle Q\rangle) \quad \text{by the definition of } \alpha^* \\
&= \alpha^* \left(\sum_{Q \in E} \operatorname{ord}_Q(r)\langle Q\rangle \right) \\
&= \alpha^*(\operatorname{div} r).
\end{aligned}$$

□

For later computations it is important to have a look at the composition of two rational maps:

Proposition 3.15 *Let α and β be non-constant rational maps. Then $\beta \circ \alpha$ is non-constant, and the following formulae hold:*

$$\begin{aligned}
e_{\beta\circ\alpha}(P) &= e_\alpha(P)\, e_\beta(\alpha(P)) \quad \forall P \in E \\
(\beta \circ \alpha)^* &= \alpha^* \circ \beta^*
\end{aligned}$$

Proof: Since α and β are surjective, so is $\beta \circ \alpha$ which thus cannot be constant. The rest of the proof consists of straightforward computations. Let u be a uniformising parameter at $(\beta \circ \alpha)(P)$.

$$\begin{aligned}
e_{\beta\circ\alpha}(P) &= \operatorname{ord}_P((u \circ \beta) \circ \alpha) \quad \text{by definition} \\
&= e_\alpha(P) \operatorname{ord}_{\alpha(P)}(u \circ \beta) \quad \text{by Lemma 3.14} \\
&= e_\alpha(P)\, e_\beta(\alpha(P)) \quad \text{by a second application of Lemma 3.14}
\end{aligned}$$

$$\begin{aligned}
(\beta \circ \alpha)^*(\langle Q \rangle) &= \sum_{P \in (\beta \circ \alpha)^{-1}(Q)} e_{\beta \circ \alpha}(P)\langle P \rangle \quad \text{by definition} \\
&= \sum_{P \in (\beta \circ \alpha)^{-1}(Q)} e_\beta(\alpha(P)) e_\alpha(P)\langle P \rangle \quad \text{by the previous formula} \\
&= \sum_{R \in \beta^{-1}(Q)} e_\beta(R) \sum_{P \in \alpha^{-1}(R)} e_\alpha(P)\langle P \rangle \\
&= \sum_{R \in \beta^{-1}(Q)} e_\beta(R) \alpha^*(\langle R \rangle) \\
&= \alpha^* \left(\sum_{R \in \beta^{-1}(Q)} e_\beta(R)\langle R \rangle \right) \\
&= \alpha^* \circ \beta^*(\langle Q \rangle)
\end{aligned}$$

□

A first example is given by the translation maps.

Lemma 3.16 *For $Q \in E$ the map τ_Q is unramified.*

Proof: Note that τ_Q has the inverse rational map τ_{-Q}. It follows that

$$1 = e_{\mathrm{id}}(P) = e_{\tau_{-Q} \circ \tau_Q}(P) = e_{\tau_Q}(P)\, e_{\tau_{-Q}}(P + Q).$$

Both of $e_{\tau_Q}(P)$ and $e_{\tau_{-Q}}(P + Q)$ are at least 1, so they are exactly 1. □

This observation about translations can be used to prove the following result on endomorphisms, for which the ramification is particularly simple.

Theorem 3.17 *Let α be a non-zero endomorphism. Then $e_\alpha = e_\alpha(P)$ is independent of the point P.*

Proof: Let P be a point on E. Since $\alpha(Q + P) = \alpha(Q) + \alpha(P)$ for all $Q \in E$ we have $\alpha \circ \tau_P = \tau_{\alpha(P)} \circ \alpha$. Now

$$\begin{aligned}
e_\alpha(P) &= \frac{e_{\alpha \circ \tau_P}(\mathcal{O})}{e_{\tau_P}(\mathcal{O})} \quad \text{by Proposition 3.15} \\
&= e_{\alpha \circ \tau_P}(\mathcal{O}) \quad \text{by Lemma 3.16} \\
&= e_{\tau_{\alpha(P)} \circ \alpha}(\mathcal{O}) \\
&= e_\alpha(\mathcal{O})\, e_{\tau_{\alpha(P)}}(\alpha(\mathcal{O})) \\
&= e_\alpha(\mathcal{O}).
\end{aligned}$$

□

Corollary 3.18 *Let α and β be non-zero endomorphisms. Then*

$$e_{\beta \circ \alpha} = e_\alpha\, e_\beta.$$

Proof: The assertion follows directly from the theorem and Proposition 3.15. □

As the first non-trivial example we examine the Frobenius endomorphism φ for $k = \mathbb{F}_q$:

Proposition 3.19

$$e_\varphi = q$$

Proof: Recall from Section 2.7 that $\frac{X}{Y}$ is a uniformising parameter at $\mathcal{O} = \varphi(\mathcal{O})$. Then

$$e_\varphi = e_\varphi(\mathcal{O}) = \operatorname{ord}_{\mathcal{O}}\left(\frac{X}{Y} \circ \varphi\right) = \operatorname{ord}_{\mathcal{O}}\left(\left(\frac{X}{Y}\right)^q\right) = q.$$

□

The second notion to be introduced in this section is that of the *degree* of an endomorphism α. The image of α^* is a subfield of $K(E)$, so a natural definition of $\deg \alpha$ would be the degree of the field extension $[K(E) : \alpha^*(K(E))]$. Since $K(E)$ and $\alpha^*(K(E))$ are both extensions of K of transcendence degree one and $[K(E) : K(X)] < \infty$, it follows that $\deg \alpha$ is finite. It can be shown that $\deg \alpha = e_\alpha \; |\ker \alpha|$ (see [Shafarevich, 1974], pp. 141–143); the proof is the exact analogue of the theorem on number fields relating the ramification and the inertia indices to the degree of the field extension. Furthermore, e_α is the inseparable, $|\ker \alpha|$ the separable degree of the field extension. Since the proofs are quite involved we use the theorem as a definition and have to accept the drawback that we loose important information and are hardly able to make use of the degree throughout this book.

Definition 3.20 *The* degree *of a non-zero endomorphism* α *is*

$$\deg \alpha = e_\alpha \; |\ker \alpha|.$$

Example. We have seen in Proposition 3.19 that $e_\varphi = q$. Moreover $\ker \varphi = \{\mathcal{O}\}$ because for a finite point $P = (x, y)$ its image $\varphi(P) = (x^q, y^q)$ is finite again. So $\deg \varphi = q$. Write $J = \varphi^*(K(E))$, and let us have a look at the field extension $K(E)/J$. Since $K(E)$ is generated over K by X and Y and φ^* fixes K, J is generated over K by X^q and Y^q, so it is the quotient field of $K[X^q, Y^q]/(E)$. Then $K(E) = J(X, Y)$. For $p \neq 2$ we can assume E in the normal form $Y^2 = X^3 + a_2X^2 + a_4X + a_6$ so that

$$Y = \frac{(Y^2)^{\frac{q+1}{2}}}{Y^q} = \frac{(X^3 + a_2X^2 + a_4X + a_6)^{\frac{q+1}{2}}}{Y^q} \in J(X)$$

and $K(E) = J(X)$. If $q = 2^m$, then

$$Y = \frac{X^3 + a_2X^2 + a_4X + a_6 + Y^2}{a_1X + a_3} \in J(X, Y^2),$$

implying that $K(E) = J(X,Y) = J(X,Y^2) = J(X,Y^4) = \cdots = J(X,Y^q) = J(X)$. Note that X is a root of the irreducible polynomial $T^q - X^q \in J[T]$, so the degree $[K(E) : J]$ is indeed $q = \deg \varphi$. Moreover the polynomial is inseparable, so $K(E)/J$ is purely inseparable, which corresponds to the observation that $|\ker \varphi| = 1$.

Proposition 3.21 *Let $\Delta \in \mathrm{Div}(E)$, α and β non-zero endomorphisms.*

1. $\deg(\alpha^*(\Delta)) = \deg \alpha \, \deg \Delta$

2. $\deg(\alpha \circ \beta) = \deg \alpha \, \deg \beta$

Proof:

1. By the linearity of the degree function on divisors and of α^* it is sufficient to consider the case $\Delta = \langle Q \rangle$.

$$\begin{aligned} \deg(\alpha^*(\Delta)) &= \deg\left(e_\alpha \sum_{P \in \alpha^{-1}(Q)} \langle P \rangle \right) \\ &= e_\alpha \, |\alpha^{-1}(Q)| \\ &= e_\alpha \, |\ker \alpha| \\ &= \deg \alpha \end{aligned}$$

2.

$$\begin{aligned} \deg(\alpha \circ \beta) &= e_{\alpha\circ\beta} |\ker(\alpha \circ \beta)| \\ &= e_\alpha \, e_\beta \, |\{P \in E : \beta(P) \in \ker \alpha\}| \quad \text{by Corollary 3.18} \\ &= e_\alpha \, e_\beta \, |\ker \alpha| \, |\ker \beta| \\ &= \deg \alpha \, \deg \beta \end{aligned}$$

□

Our aim is to investigate the maps $[m]$ and the m-torsion points $E[m]$. In particular we are interested in the ramification index $e_{[m]}$. This requires some preparations, which are done during the following sections.

3.3 A DERIVATION ON $K(E)$

In this section we define a derivation on $K(E)$ which plays the same role as the usual derivation on $K(X)$ in the sense that it can be used to investigate the multiplicities of zeros of rational functions.

Definition 3.22 *Let L be a K-algebra. A* derivation *on L is a K-linear map $D : L \to L$ which satisfies the product rule*

$$D(fg) = f\,Dg + g\,Df.$$

Applying the product rule to $f = g = 1$ it follows that $D1 = 2\,D1$, so $D1 = 0$ and $Dc = c\,D1 = 0$ for constants $c \in K$. If g is a unit in L, then letting $f = \frac{1}{g}$ implies

$$0 = D\left(g\frac{1}{g}\right) = g\,D\frac{1}{g} + \frac{Dg}{g},$$

so

$$D\frac{1}{g} = -\frac{Dg}{g^2}.$$

Another application of the product rule to $D\left(f\frac{1}{g}\right)$ yields the quotient rule

$$D\frac{f}{g} = \frac{g\,Df - f\,Dg}{g^2}.$$

So the derivatives of cf, $f \pm g$, fg and $\frac{f}{g}$ for $f \in L$, $g \in L^\times$ and $c \in K$ are uniquely determined by the values of Df and Dg.

We first determine all possible derivations on $K(X,Y)$.

Proposition 3.23 *For $f, g \in K(X,Y)$, there is a unique derivation D on $K(X,Y)$ such that $DX = f$ and $DY = g$, which is given by*

$$Dr = \left(\frac{\partial r}{\partial X}, \frac{\partial r}{\partial Y}\right)\begin{pmatrix} f \\ g \end{pmatrix} \quad \textit{for } r \in K(X,Y).$$

Proof: For the uniqueness note that $K[X,Y]$ is generated as a K-algebra by X and Y, so the values of DX and DY determine D on $K[X,Y]$, and the quotient rule assures that there is a unique extension of D to $K(X,Y)$.

From a theoretical point of view the existence of a derivation with prescribed values DX and DY follows from the algebraic independence of X and Y over K. For the actual construction consider the K-linear function

$$D' : K(X,Y) \to K(X,Y), \quad r \mapsto \left(\frac{\partial r}{\partial X}, \frac{\partial r}{\partial Y}\right)\begin{pmatrix} f \\ g \end{pmatrix}.$$

D' defines a derivation on $K(X,Y)$: The linearity and the validity of the product rule follow directly from the linearity and the product rule for partial derivatives. Moreover $D'X = f$ and $D'Y = g$, which provides a constructive proof of the existence. □

When defining a derivation on $K(E)$, the field of fractions of $K[X,Y]/(E)$, we must assure that $DE = D0 = 0$. So we cannot choose DX and DY arbitrarily any more, but DY depends on DX.

Proposition 3.24 *For $f \in K(E)$ there is a unique derivation D on $K(E)$ such that $DX = f$. It is given by*

$$DY = \frac{3X^2 + 2a_2X + a_4 - a_1Y}{2Y + a_1X + a_3}DX$$

and

$$Dr = \left(\frac{\partial r}{\partial X}, \frac{\partial r}{\partial Y}\right)\begin{pmatrix} DX \\ DY \end{pmatrix} \quad \textit{for } r \in K(E).$$

Proof: For the uniqueness note that $0 = E$ in $K(E)$, and using the product rule one computes

$$0 = DE = (2Y + a_1X + a_3)DY - (3X^2 + 2a_2X + a_4 - a_1Y)DX,$$

so DY is uniquely determined by DX. Then as in the previous proposition there is at most one way to extend D to $K[E]$ since $K[E]$ is generated by X and Y. Finally the quotient rule determines D on the whole of $K(E)$.

For the proof of the existence consider the function

$$D : K(E) \to K(E), \quad r \mapsto \left(\frac{\partial r}{\partial X}, \frac{\partial r}{\partial Y}\right)\begin{pmatrix} DX \\ DY \end{pmatrix},$$

where DX and DY are as in the proposition. It follows that $DE = 0$, so D is well defined on $K(E)$. Again it is easy to verify that D indeed defines a derivation. □

The two propositions above can easily be generalised to arbitrary function fields: Suppose that L is a function field over K of transcendence degree n, i.e. there are algebraically independent elements $X_1, \ldots, X_n \in L$ such that $L/K(X_1, \ldots, X_n)$ is separable. Then for $f_1, \ldots, f_n \in L$ there is a unique derivation D on L such that $DX_i = f_i$ for $1 \leq i \leq n$.

The close relationship between a derivation on $K(E)$ and partial derivatives can be exploited to investigate the derivative of a rational function which is composed with a rational map.

Proposition 3.25 *Let $\alpha = (\alpha_1, \alpha_2)$ be a rational map and r a rational function. Then*

$$D(r \circ \alpha) = \left(\frac{\partial r}{\partial X} \circ \alpha, \frac{\partial r}{\partial Y} \circ \alpha\right)\begin{pmatrix} \frac{\partial \alpha_1}{\partial X} & \frac{\partial \alpha_1}{\partial Y} \\ \frac{\partial \alpha_2}{\partial X} & \frac{\partial \alpha_2}{\partial Y} \end{pmatrix}\begin{pmatrix} DX \\ DY \end{pmatrix}$$

Proof: The assertion follows easily from the previous proposition and the chain rule for the partial derivatives of $r \circ \alpha$; we renounce at details. □

In view of Proposition 3.24 there is a natural choice for DX, which assures that $DY \in K[E]$ so that the restriction of D to $K[E]$ defines a derivation as well.

Definition 3.26 *The canonical derivation on $K(E)$ is given by*

$$\begin{aligned} DX &= 2Y + a_1X + a_3 = \frac{\partial E}{\partial Y} \\ DY &= 3X^2 + 2a_2X + a_4 - a_1Y = -\frac{\partial E}{\partial X} \\ Dr &= \left(\frac{\partial r}{\partial X}, \frac{\partial r}{\partial Y}\right)\begin{pmatrix} DX \\ DY \end{pmatrix} \quad \textit{for } r \in K(E). \end{aligned}$$

We fix this derivation during the rest of this section. As an important example we compute the derivatives of g_m and h_m.

Theorem 3.27 *Let m be a positive integer. Then the following identities hold:*

$$\begin{aligned}
Dg_m &= m(2h_m + a_1 g_m + a_3) \\
&= m\frac{\partial E}{\partial Y} \circ [m] \\
Dh_m &= m(3g_m^2 + 2a_2 g_m + a_4 - a_1 h_m) \\
&= -m\frac{\partial E}{\partial X} \circ [m]
\end{aligned}$$

Proof: We use induction on m, treating the cases $m = 1$ and $m = 2$ separately. As the computations are rather tedious, not all the details are given. The reader is invited to check the correctness of the results with the help of a symbolic algebra programme.

The case $m = 1$ is just the definition of DX and DY.

In the case $m = 2$ we use the duplication formula to express [2] and obtain the following relations:

$$\begin{aligned}
DX &= 2Y + a_1X + a_3, \quad DY = 3X^2 + 2a_2X + a_4 - a_1Y && (3.4) \\
\lambda &= \frac{3X^2 + 2a_2X + a_4 - a_1Y}{2Y + a_1X + a_3} && (3.5) \\
g_2 &= -2X + \lambda^2 + a_1\lambda - a_2 && (3.6) \\
h_2 &= -(\lambda + a_1)g_2 - a_3 - Y + \lambda X && (3.7)
\end{aligned}$$

Then the derivatives are as follows:

$$\begin{aligned}
D\lambda &= \frac{((6X + 2a_2)DX - a_1DY)(2Y + a_1X + a_3)}{(2Y + a_1X + a_3)^2} && (3.8) \\
&\quad - \frac{(3X^2 + 2a_2X + a_4 - a_1Y)(2DY + a_1DX)}{(2Y + a_1X + a_3)^2} \\
Dg_2 &= -2DX + 2\lambda D\lambda + a_1D\lambda && (3.9) \\
Dh_2 &= -(\lambda + a_1)Dg_2 - g_2D\lambda - DY + \lambda DX + XD\lambda && (3.10)
\end{aligned}$$

Plugging DX and DY from (3.4) into (3.8) to (3.10), the expression for $D\lambda$ into (3.9) and (3.10) and the resulting expression for Dg_2 into (3.10) we obtain Dg_2 and Dh_2 as rational functions in X and Y. Similarly we use (3.5) to (3.7) to express g_2 and h_2 in terms of X and Y. We form $Dg_2 - 2(2h_2 + a_1g_2 + a_3)$ and $Dh_2 - 2(3g_2^2 + 2a_2g_2 + a_4 - a_1h_2)$, multiply everything out and eliminate higher powers of Y. Then both expressions are zero.

For $m > 2$ we use the generic addition formula (see Corollary 3.10) and obtain the following relations:

$$\begin{aligned}
\lambda &= \frac{h_{m-1} - Y}{g_{m-1} - X} && (3.11)\\
g_m &= -g_{m-1} - X + \lambda^2 + a_1\lambda - a_2 && (3.12)\\
h_m &= -(\lambda + a_1)g_m - a_3 - Y + \lambda X && (3.13)\\
D\lambda &= \frac{(Dh_{m-1} - DY)(g_{m-1} - X)}{(g_{m-1} - X)^2} && (3.14)\\
&\quad - \frac{(h_{m-1} - Y)(Dg_{m-1} - DX)}{(g_{m-1} - X)^2} &&\\
Dg_m &= -Dg_{m-1} - DX + 2\lambda D\lambda + a_1 D\lambda && (3.15)\\
Dh_m &= -(\lambda + a_1)Dg_m - g_m D\lambda - DY + \lambda DX + XD\lambda && (3.16)
\end{aligned}$$

By the induction hypothesis we have furthermore

$$\begin{aligned}
Dg_{m-1} &= (m-1)(2h_{m-1} + a_1 g_{m-1} + a_3) && (3.17)\\
Dh_{m-1} &= (m-1)(3g_{m-1}^2 + 2a_2 g_{m-1} + a_4 - a_1 h_{m-1}) && (3.18)
\end{aligned}$$

Plugging in (3.4) and the identities above we can express $Dg_m - m(2h_m + a_1 g_m + a_3)$ and $Dh_m - m(3g_m^2 + 2a_2 g_m + a_4 - a_1 h_m)$ in terms of X, Y, g_{m-1} and h_{m-1}. Noting that (X, Y) and (g_{m-1}, h_{m-1}) are rational maps we eliminate occurrences of Y^2 and h_{m-1}^2 and find again that the expressions are zero. □

For polynomials $f \in K[X]$ it is well known that $f' = 0$ if and only if $f = f_1(X^p)$ for a suitable polynomial f_1. We would like to obtain a similar result for rational functions.

Lemma 3.28 *Let $v \in K(X)$ be a rational function in one variable. Then*

$$v' = 0 \Leftrightarrow v = v_1(X^p) \text{ for a suitable } v_1 \in K(X).$$

Proof: It is clear that

$$\frac{\partial v_1(X^p)}{\partial X} = pX^{p-1}v_1'(X^p) = 0$$

in characteristic p. To see the converse, suppose that $v = \frac{f}{g}$ with coprime $f, g \in K[X]$. Then

$$0 = v' = \frac{f'g - fg'}{g^2}$$

implies that

$$f'g = fg',$$

so $f|f'$ and $g|g'$. As $\deg f' < \deg f$ (unless $f = 0$, in which case nothing is to show) and $\deg g' < \deg g$, we conclude that $f' = g' = 0$, so $f = f_1(X^p)$ and $g = g_1(X^p)$ for suitable polynomials $f_1, g_1 \in K[X]$, and $v = \frac{f_1}{g_1}(X^p)$. □

Theorem 3.29 *Let $r \in K(E)$, $p > 0$. Then*

$$Dr = 0 \Leftrightarrow r = r_1(X^p, Y^p) \textit{ for a suitable } r_1 \in K(E).$$

Proof: $Dr_1(X^p, Y^p) = 0$ follows from Proposition 3.25. For the converse implication, observe first that $Y^p \notin K(X)$: If $p = 2$, we have $Y^2 = (a_1 X + a_3)Y + (X^3 + a_2 X^2 + a_4 X + a_6)$ with $a_1 X + a_3 \neq 0$. Then $Y \notin K(X)$ implies $Y^2 \notin K(X)$. If $p \neq 2$, then Y^p is a root of the polynomial

$$T^2 + a_1^p X^p T + a_3^p T - (X^{3p} + a_2^p X^{2p} + a_4^p X^p + a_6^p) \in K(X)[T],$$

which can be shown to be irreducible over $K(X)$ with the same argumentation as in Section 2.2. (There the crucial point of the proof was that $\deg f + \deg g = 3$ was odd; here $\deg f + \deg g = 3p$ is odd again.)

So $K(X) \subsetneq K(X, Y^p) \subseteq K(E)$, and

$$2 = [K(E) : K(X)] = [K(E) : K(X, Y^p)][K(X, Y^p) : K(X)]$$

implies that $K(X, Y^p) = K(E)$ and that $\{1, Y^p\}$ is a basis of $K(E)$ over $K(X)$. So $r \in K(E)$ has a representation $r = u + Y^p v$ with $u, v \in K(X)$. Moreover, write $Y^p = f + gY$ with $f, g \in K(X)$, $g \neq 0$; let $t = a_1 X + a_3$ and $s = X^3 + a_2 X^2 + a_4 X + a_6$. Then

$$Dr = (u' + Y^p v')DX + pY^{p-1} v DY = (u' + Y^p v')DX.$$

$Dr = 0$ and $DX \neq 0$ imply $u' + Y^p v' = 0$. Since $\{1, Y^p\}$ is a basis, this in turn yields $u' = v' = 0$. Using the previous lemma, $u = u_1(X^p)$ and $v = v_1(X^p)$, so $r = r_1(X^p, Y^p)$ with $r_1 = u_1 + Y v_1$. □

We now try to compare the multiplicities of zeros and poles of a rational function and its derivative. For a polynomial f in characteristic zero it is known that if f has a zero of order $d > 0$ in x, then f' has a zero of order $d - 1$ in x. This still holds for rational functions with the usual precaution concerning the characteristic.

Theorem 3.30 *Let r be a rational function, $P \in E$ and $d = \mathrm{ord}_P(r)$.*

- *If $p \nmid d$, then $\mathrm{ord}_P(Dr) = d - 1$.*
- *If $p|d$, then $\mathrm{ord}_P(Dr) \geq d$.*

Proof: By convention, for $p = 0$ the first case applies to $d > 0$, the second case to $d = 0$. We first prove the special case $d = 0$. If $P \neq \mathcal{O}$, write $r = \frac{f}{g}$ with $f(P), g(P) \neq 0$. Then $Dr = \frac{gDf - fDg}{g^2}$ is regular at P. For $P = \mathcal{O}$, write $r = u + vY$ with $u, v \in K(X)$ and recall from Lemma 2.31 that

$0 = d = \mathrm{ord}_{\mathcal{O}}\, r = \min\{-2\deg u, -3 - 2\deg v\}$. It follows that $\deg u = 0$, so u is a constant and $Du = 0$, and that $-3 - 2\deg v \geq 0$ or equivalently $-2\deg v \geq 4$. From $\deg v' \leq \deg v - 1$ we deduce that $-2\deg v' \geq -2\deg v + 2 \geq 6$. We compute Dr as

$$\begin{aligned} Dr &= Du + v'YDX + vDY \\ &= v'Y(2Y + a_1X + a_3) + v(3X^2 + 2a_2X + a_4 - a_1Y). \end{aligned}$$

Note that $\mathrm{ord}_{\mathcal{O}}(2Y + a_1X + a_3) \geq -3$ and $\mathrm{ord}_{\mathcal{O}}(3X^2 + 2a_2X + a_4 - a_1Y) \geq -4$ by Lemma 2.31. Hence,

$$\begin{aligned} \mathrm{ord}_{\mathcal{O}}(Dr) &\geq \min\{6 - 3 + \mathrm{ord}_{\mathcal{O}}(2Y + a_1X + a_3), \\ &\qquad 4 + \mathrm{ord}_{\mathcal{O}}(3X^2 + 2a_2X + a_4 - a_1Y)\} \\ &\geq 0. \end{aligned}$$

The next step is the proof for $d = 1$. Suppose that u is the uniformising parameter of the Sections 2.5 and 2.7 at P, such that $r = u\varepsilon$ with $\mathrm{ord}_P\, \varepsilon = 0$. Then

$$\mathrm{ord}_P(Dr) = \mathrm{ord}_P(\varepsilon\, Du + u\, D\varepsilon).$$

If we can show that $\mathrm{ord}_P(Du) = 0$, then $\mathrm{ord}_P(Dr) = 0$ because $\mathrm{ord}_P(\varepsilon\, Du) = \mathrm{ord}_P\, \varepsilon + \mathrm{ord}_P(Du) = 0$ and $\mathrm{ord}_P(u\, D\varepsilon) = 1 + \mathrm{ord}_P(D\varepsilon) \geq 1$ by the case $d = 0$. We proceed to prove that $\mathrm{ord}_P(Du) = 0$ by examining the different cases:

- $P = (x, y) \notin E[2]$, i.e. $2y + a_1x + a_3 \neq 0$; $u = X - x$
  Then $Du = DX = 2Y + a_1X + a_3$ is regular and not zero in P.

- P is of order 2 and $p \neq 2$; $u = 2Y + a_1X + a_3 = \dfrac{\partial E}{\partial Y}$
  Thus $\dfrac{\partial E}{\partial Y}(P) = 0$ and $\dfrac{\partial E}{\partial X}(P) \neq 0$ by the non-singularity of E, so $Du(P) = 2DY(P) + a_1DX(P) = 2\dfrac{\partial E}{\partial X}(P) + a_1\dfrac{\partial E}{\partial Y}(P) \neq 0$.

- P is of order 2 and $p = 2$; $u = Y + y$
  As in the previous case $\dfrac{\partial E}{\partial Y}(P) = 0$ and $\dfrac{\partial E}{\partial X}(P) \neq 0$, hence $Du(P) = DY = \dfrac{\partial E}{\partial X}(P) \neq 0$.

- $P = \mathcal{O}$; $u = \frac{X}{Y}$

$$\begin{aligned} Du &= \frac{YDX - XDY}{Y^2} \\ &= \frac{Y(2Y + a_1X + a_3) - X(3X^2 + 2a_2X + a_4 - a_1Y)}{Y^2} \\ &= \frac{-X^3 + a_4X + 2a_6 - a_3Y}{X^3 + a_2X^2 + a_4X + a_6 - (a_1X + a_3)Y} \\ \mathrm{ord}_{\mathcal{O}}(Du) &= -6 - (-6) = 0 \quad \text{by Lemma 2.31} \end{aligned}$$

Finally let $d \geq 2$. Write $r = u^d r_1$ with u a uniformising parameter at P and $\operatorname{ord}_P r_1 = 0$. Then

$$\begin{aligned} Dr &= du^{d-1} r_1 Du + u^d Dr_1 \\ &= u^{d-1}(dr_1 Du + uDr_1) \end{aligned}$$

where $\operatorname{ord}_P(uDr_1) \geq 1$ and $\operatorname{ord}_P(r_1 Du) = 0$. If $p \nmid d$, we deduce that $\operatorname{ord}_P(Dr) = d - 1$; if $p|d$, then $Dr = u^d Dr_1$, and $\operatorname{ord}_P(Dr) \geq d$. □

We finish this section with the observation that derivation and multiplication by m "almost commute":

Proposition 3.31 *Let m be an integer and r a rational function. Then*

$$D(r \circ [m]) = m\, Dr \circ [m].$$

Proof: It is easy to see that the set of rational functions which satisfy the proposition is closed under the field operations. Thus it is sufficient to consider the cases $r = X$ and $r = Y$. If $m > 0$, this is just Theorem 3.27. If $m = 0$, it is trivial. For $m = -1$ we perform some basic computations:

$$\begin{aligned} D(X \circ [-1]) &= DX \\ &= 2Y + a_1 X + a_3 \\ &= -(2(-Y - a_1 X - a_3) + a_1 X + a_3) \\ &= -DX \circ [-1] \\ D(Y \circ [-1]) &= D(-Y - a_1 X - a_3) \\ &= -((3X^2 + 2a_2 X + a_4) - a_1(-Y - a_1 X - a_3)) \\ &= -DY \circ [-1] \end{aligned}$$

For $m < -1$ we combine the information we already have:

$$\begin{aligned} D(r \circ [m]) &= D(r \circ [-m] \circ [-1]) \\ &= -D(r \circ [-m]) \circ [-1] \\ &= -(-m)\, Dr \circ [-m] \circ [-1] \\ &= m\, Dr \circ [m] \end{aligned}$$

□

3.4 SEPARABILITY

Among other results we determine $e_{[m]}$ and the ramification indices of an important class of endomorphisms in this section.

THROUGHOUT THE SECTION WE ASSUME $p > 0$.

Definition 3.32 *A non-zero endomorphism α is called* separable *if $e_\alpha = 1$,* inseparable *otherwise.*

In fact the definition of "separable" coincides with that of "unramified" in Section 3.2. It is motivated by the fact that for a separable endomorphism α the field extension $K(E)/\alpha^*(K(E))$ is separable, see the remark at the end of Section 3.2.

Proposition 3.33 *The following assertions are equivalent for a non-zero endomorphism α:*

1. *α is inseparable.*
2. $D(r \circ \alpha) = 0 \quad \forall r \in K(E)$
3. *$\alpha = (r(X^p, Y^p), s(X^p, Y^p))$ for suitable rational functions r and s.*

Proof:

1. $\Rightarrow$ 2.
Suppose α is inseparable, r a rational function with $D(r \circ \alpha) \neq 0$. Then $D(r \circ \alpha)$ has only a finite number of zeros and poles, so there must be a point $P \in E$ in which $D(r \circ \alpha)$ is defined and not zero. Let $s := r - (r \circ \alpha)(P)$. Then $s \circ \alpha(P) = 0$, so $\text{ord}_P(s \circ \alpha) \geq 1$. On the other hand $D(s \circ \alpha)(P) = D(r \circ \alpha)(P) \neq 0$ so that $\text{ord}_P(D(s \circ \alpha)) = 0$ and $\text{ord}_P(s \circ \alpha) = 1$ by Theorem 3.30. But by Lemma 3.14, $\text{ord}_P(s \circ \alpha) = e_\alpha \, \text{ord}_{\alpha(P)}(s) > 1$ since $e_\alpha > 1$, a contradiction.

2. $\Rightarrow$ 3.
Letting $r = X$ and $r = Y$ this is a direct consequence of Theorem 3.29.

3. $\Rightarrow$ 1.
By definition of e_α and since $\frac{X}{Y}$ is a uniformising parameter in $\mathcal{O}$ (see Theorem 2.26),

$$e_\alpha = \text{ord}_{\mathcal{O}}\left(\frac{X}{Y} \circ \alpha\right) = \text{ord}_{\mathcal{O}}\left(\frac{r(X^p, Y^p)}{s(X^p, Y^p)}\right).$$

We compute $D\left(\frac{X}{Y} \circ \alpha\right)$:

$$D\left(\frac{X}{Y} \circ \alpha\right) = \frac{s(X^p, Y^p) D(r(X^p, Y^p)) - r(X^p, Y^p) D(s(X^p, Y^p))}{s(X^p, Y^p)^2}$$

Note that

$$\begin{aligned} D(r(X^p, Y^p)) &= \frac{\partial(r(X^p, Y^p))}{\partial X} DX + \frac{\partial(r(X^p, Y^p))}{\partial Y} DY \\ &\quad \text{by Proposition 3.24} \\ &= 0 \end{aligned}$$

and $D(s(X^p, Y^p)) = 0$ by the same argument. So

$$D\left(\frac{X}{Y} \circ \alpha\right) = 0 \text{ and } \text{ord}_{\mathcal{O}}\left(D\left(\frac{X}{Y} \circ \alpha\right)\right) \geq 1.$$

By Theorem 3.30, it follows that

$$\operatorname{ord}_{\mathcal{O}}\left(\frac{X}{Y}\circ\alpha\right) > 1,$$

and α is inseparable.

□

Corollary 3.34 *Let α and β be non-zero endomorphisms.*

1. If α and β are inseparable, then so is $\alpha+\beta$.

2. If α is separable and β inseparable, then $\alpha+\beta$ is separable

Proof:

1. This is an immediate consequence of Proposition 3.33, 3.
2. The assertion follows from 1.: If $\alpha+\beta$ were inseparable, then so would be $\alpha=(\alpha+\beta)-\beta$, a contradiction.

□

Proposition 3.35 *Let m be coprime to p. Then $[m]$ is separable.*

Proof: Let $P \in E$ and u a uniformising parameter at mP, such that $e_{[m]} = \operatorname{ord}_P(u\circ[m])$. We have seen in Proposition 3.31 that

$$D(u\circ[m]) = m\,Du\circ[m].$$

It follows that

$$\begin{aligned}\operatorname{ord}_P(D(u\circ[m])) &= \operatorname{ord}_P(Du\circ[m])\\ &= e_{[m]}\operatorname{ord}_{mP}(Du) \quad \text{by Lemma 3.14}\\ &= 0 \quad \text{by Theorem 3.30.}\end{aligned}$$

Another application of Theorem 3.30 yields $\operatorname{ord}_P(u\circ[m]) = 1$ as desired. □

Corollary 3.36 *If $k=\mathbb{F}_q$ and m and n are integers with m coprime to p, then $[m]+[n]\circ\varphi$ is separable.*

Proof: By Corollary 3.18 and Proposition 3.19 we have $e_{[n]\circ\varphi} = e_{[n]}\,e_\varphi = qe_{[n]} > 1$, so $[n]\circ\varphi$ is inseparable. Since $[m]$ is separable the conclusion is immediate from Corollary 3.34, 2. □

3.5 m-TORSION POINTS

The topic of this section is the group structure of $E[m]$. Precisely, we show the following results:

Proposition 3.37 *If m is coprime to p, then*

$$E[m] \simeq \mathbb{Z}_m \times \mathbb{Z}_m.$$

Proposition 3.38 *If $E[p] \neq \{\mathcal{O}\}$, then*

$$E[p^\nu] \simeq \mathbb{Z}_{p^\nu}.$$

These two results can be combined, using elementary group theory, to yield the following theorem.

Theorem 3.39 *Let m be a positive integer, $m = p^\nu m'$ where $p \nmid m'$.*

- *If $E[p] = \{\mathcal{O}\}$, then*

$$E[m] \simeq \mathbb{Z}_{m'} \times \mathbb{Z}_{m'}.$$

- *Otherwise*

$$E[m] \simeq \mathbb{Z}_{m'} \times \mathbb{Z}_m.$$

We obtain these results by examining g_m and h_m and the divisor of $g_m - g_n$. It is perhaps in this section that we have to pay the highest price for our elementary approach to the topic: We need a lot of explicit computations, always treating the cases of characteristic 2 and 3 separately. Although the computations are feasible by hand, a symbolic algebra programme provides a considerable support. To ease the presentation we introduce the following notion:

Definition 3.40 *The* leading coefficient *(in $\mathcal{O}$) of a rational function r is*

$$l(r) := \left(\left(\frac{X}{Y}\right)^{-\operatorname{ord}_{\mathcal{O}} r} r\right)(\mathcal{O})$$

l is a multiplicative homomorphism from $K(E)^\times$ to $K^\times$. In addition to Proposition 2.14 we have for two rational functions r and s with the same order in $\mathcal{O}$ that this order is preserved in $r+s$ if and only if the leading coefficients of r and s do not sum to zero. In this case $l(r+s) = l(r) + l(s)$. Since the leading coefficients of X and Y are 1, we can multiply a rational function by powers of X and Y without changing its leading coefficient.

We need some information on the order and leading coefficient of g_m in $\mathcal{O}$. As usual our proofs are by induction on m with suitable modifications since most of the results are only valid for m coprime to the characteristic p. As a preparation we examine g_p for $p \in \{2,3\}$; a first useful result is the number of p-torsion points:

Proposition 3.41 *Let $p \in \{2,3\}$. Then*

$$|E[p]| = \begin{cases} p & \textit{if } j \neq 0 \\ 1 & \textit{if } j = 0 \end{cases}$$

Proof: The result is well-known for $p = 2$, see Section 2.5. For $p = 3$ note that a finite point $P \in E$ lies in $E[3]$ if and only if $2P = \pm P$, which means that $X(2P) = X(P)$ or $(g_2 - X)(P) = 0$. We compute $g_2 - X$, and eliminating higher powers of Y in the numerator we find that

$$g_2 - X = \frac{-b_2 X^3 - b_8}{(-Y + a_1 X + a_3)^2},$$

where b_2 and b_8 are as in Definition 2.7. Since we are in characteristic 3, for $b_2 \neq 0$ or equivalently $j = \frac{b_2^6}{\Delta} \neq 0$ the numerator has the unique root $x = \sqrt[3]{-\frac{b_8}{b_2}}$, to which correspond two points $P = (x, y)$ and $Q = \overline{P}$ on E, where y is a root of $E(x, Y)$. By Lemma 2.31 the numerator has a pole of order 6 in $\mathcal{O}$, so its divisor must be $3\langle P\rangle + 3\langle Q\rangle - 6\langle\mathcal{O}\rangle$. The denominator has the divisor $2\langle R_1\rangle + 2\langle R_2\rangle + 2\langle R_3\rangle - 6\langle\mathcal{O}\rangle$ where the R_i's are the three distinct points of order 2. It follows that $\mathrm{ord}_P(g_2 - X)$, $\mathrm{ord}_Q(g_2 - X) \geq 1$, and $E[3] = \{P, Q, \mathcal{O}\}$. On the other hand, for $j = 0$ or equivalently $b_2 = 0$, $g_2 - X$ has no zero and hence $E[3] = \{\mathcal{O}\}$. □

Proposition 3.42 *Let $p \in \{2,3\}$. Define α by*

$$\alpha = \frac{p^2}{|E[p]|} = \begin{cases} p & \textit{for } j \neq 0 \\ p^2 & \textit{for } j = 0 \end{cases}$$

Then

$$\mathrm{ord}_{\mathcal{O}}\, g_p = -2\alpha \ \textit{and}\ \mathrm{ord}_{\mathcal{O}}\, h_p = -3\alpha.$$

The leading coefficients are

$$l(g_p) = \frac{1}{\gamma^2}, \quad l(h_p) = \frac{1}{\gamma^3}$$

with

$$\gamma = \begin{cases} a_1 & \textit{for } p = 2 \textit{ and } j \neq 0 \\ a_3 & \textit{for } p = 2 \textit{ and } j = 0 \\ a_1^2 + a_2 & \textit{for } p = 3 \textit{ and } j \neq 0 \\ (a_1 a_3 - a_4)^2 & \textit{for } p = 3 \textit{ and } j = 0 \end{cases}$$

Proof:

- Assume $p = 2$, and consider the duplication formulae for $j \neq 0$ and $j = 0$. (We do not assume the curve to be in normal form.)
  - If $j \neq 0$, or equivalently $a_1 \neq 0$, the duplication formulae are
$$\begin{aligned} \lambda &= \frac{X^2 + a_4 + a_1 Y}{a_1 X + a_3} \\ g_2 &= \lambda^2 + a_1\lambda + a_2 \\ h_2 &= \lambda(g_2 + X) + Y + (a_1 g_2 + a_3) \end{aligned}$$
By Lemma 2.31, λ has a pole of order 2 in $\mathcal{O}$, and $l(\lambda) = \frac{1}{a_1}$. So g_2 has a pole of order 4 in $\mathcal{O}$ with $l(g_2) = l(\lambda^2) = \frac{1}{a_1^2}$ and h_2 has a pole of order 6 in $\mathcal{O}$ with $l(h_2) = l(\lambda)l(g_2) = \frac{1}{a_1^3}$.
  - If $j = 0$, we have $a_1 = 0$, and the duplication formulae are
$$\begin{aligned} \lambda &= \frac{X^2 + a_4}{a_3} \\ g_2 &= \lambda^2 + a_2 \\ h_2 &= \lambda(g_2 + X) + Y + a_3 \end{aligned}$$
The result can then be proved in an analogous way to the previous case.

- Now let $p = 3$. With the notations of Definition 2.7 we have $c_4 = b_2^2 = (a_1^2 + a_2)^2$, $b_4 = a_1 a_3 - a_4$ and $j = \frac{(a_1^2+a_2)^6}{\Delta}$. So $j = 0$ is equivalent to $b_2 = a_1^2 + a_2 = 0$. We first compute g_2 and h_2 by the duplication formulae:
$$\begin{aligned} \lambda_2 &= \frac{-a_2 X + a_4 - a_1 Y}{-Y + a_1 X + a_3} \\ g_2 &= X + \lambda_2^2 + a_1\lambda_2 - a_2 \\ h_2 &= -\lambda_2(g_2 - X) - Y - (a_1 g_2 + a_3) \end{aligned}$$
Since $\operatorname{ord}_{\mathcal{O}} \lambda_2 \geq 0$ by Lemma 2.31 it follows that $\operatorname{ord}_{\mathcal{O}} g_2 = \operatorname{ord}_{\mathcal{O}} X = -2$, $l(g_2) = l(X) = 1$, $\operatorname{ord}_{\mathcal{O}} h_2 = \operatorname{ord}_{\mathcal{O}}(-Y) = -3$ and $l(h_2) = -l(Y) = -1$. We then proceed to g_3 and h_3:
$$\begin{aligned} \lambda &= \frac{h_2 - Y}{g_2 - X} \\ g_3 &= \lambda^2 + (-g_2 - X + a_1\lambda - a_2) \\ h_3 &= -\lambda g_3 + (\lambda X - Y - a_1 g_3 - a_3) \end{aligned}$$
Substituting h_2 and g_2 into λ and eliminating higher powers of Y yields
$$\lambda = \frac{(X^6 + \cdots) + a_1 b_2 (X^3 + \cdots)t}{(b_2 X^3 + \cdots)t},$$

where $t = -Y + a_1X + a_3$ is of order -3 in $\mathcal{O}$ and has leading coefficient -1, and the dots stand for lower degree terms. Thus for $j \neq 0$, $\operatorname{ord}_{\mathcal{O}} \lambda = -3$ and $l(\lambda) = -\frac{1}{b_2}$. If $j = 0$, we can simplify λ to

$$\lambda = \frac{X^6 + \cdots}{b_4^2 t},$$

which has order -9 in $\mathcal{O}$ and leading coefficient $-\frac{1}{b_4^2}$. Now in both cases the term of lowest order in the expression for g_3 is λ^2, so $\operatorname{ord}_{\mathcal{O}} g_3 = 2 \operatorname{ord}_{\mathcal{O}} \lambda$ and $l(g_3) = l(\lambda)^2$, which proves the assertion for g_3. The lowest order term of h_3 is $-\lambda g_3$, such that $\operatorname{ord}_{\mathcal{O}} h_3 = 3 \operatorname{ord}_{\mathcal{O}} \lambda$ and $l(h_3) = -l(\lambda)^3$.

□

We now proceed to a result which is valid in any characteristic:

Proposition 3.43 *Let m be a positive integer coprime to p. Then g_m has a pole of order 2 and h_m has a pole of order 3 in $\mathcal{O}$, and their leading coefficients are $l(g_m) = \frac{1}{m^2}$ and $l(h_m) = \frac{1}{m^3}$.*

Proof: We use induction on m. Unfortunately, when we arrive at a multiple of p, we cannot expect to continue the induction since the proposition is not valid any more. We bridge this gap by passing from $m-1$ directly to $m+1$. To keep the overview over quite a lot of cases let us first assume that p is neither 2 nor 3. Notice that whenever we formulate a statement such as "the leading coefficients do not sum to zero" this statement has to be understood in $\mathbb{F}_p$, not in $\mathbb{Q}$ (except, of course, for $p = 0$).

1st case: $m = 1$
We already know the result for $g_1 = X$ and $h_1 = Y$, see Lemma 2.31.

2nd case: $m = 2$
We use the duplication formula to express g_2 and h_2:

$$\begin{aligned} \lambda &= \frac{3X^2 + 2a_2X + a_4 - a_1Y}{2Y + a_1X + a_3} \\ g_2 &= -2X + \lambda^2 + (a_1\lambda - a_2) \\ h_2 &= -\lambda(g_2 - X) - Y - (a_1g_2 + a_3) \end{aligned}$$

By Lemma 2.31, λ has a simple pole in $\mathcal{O}$, so $\operatorname{ord}_{\mathcal{O}}(-2X) = -2$, $\operatorname{ord}_{\mathcal{O}}(\lambda^2) = -2$ and $\operatorname{ord}_{\mathcal{O}}(a_1\lambda - a_2) \geq -1$. If we can show that the leading coefficients of $-2X$ and λ^2 sum to $\frac{1}{4}$ the assertion is proven for g_2:

$$\begin{aligned}
l(\lambda) &= \frac{\lambda X}{Y}(\mathcal{O}) \\
&= \frac{(3X^2 + 2a_2XZ + a_4Z^2 - a_1YZ)X}{(2YZ + a_1XZ + a_3Z^2)Y}(0,1,0) \\
&= \frac{(3X^2 + 2a_2XZ + a_4Z^2 - a_1Z)X}{(2Z + a_1XZ + a_3Z^2)}(0,0) \\
&= \frac{3}{2}
\end{aligned}$$

(Recall from the proof of Theorem 2.26 that X is a uniformising parameter and Z has order 3 at $(0,0)$ on E_*.)

$$l(g_2) = -2 + l(\lambda)^2 = \frac{1}{4}$$

For h_2 note that since g_2 and X have a pole of order 2 in $\mathcal{O}$ and their leading coefficients do not cancel out, we have $\mathrm{ord}_{\mathcal{O}}(g_2 - X) = -2$ and thus $\mathrm{ord}_{\mathcal{O}}(-\lambda(g_2 - X)) = -3$; furthermore $\mathrm{ord}_{\mathcal{O}}(-Y) = -3$ and $\mathrm{ord}_{\mathcal{O}}(a_1g_2 + a_3) \geq -2$. Similarly to the examination of g_2 we show that the leading coefficients of $-\lambda(g_2 - X)$ and $-Y$ sum to $\frac{1}{8}$:

$$\begin{aligned}
l(-\lambda(g_2 - X)) &= -l(\lambda)(l(g_2) - l(X)) = -\frac{3}{2}\left(\frac{1}{4} - 1\right) = \frac{9}{8} \\
l(h_2) &= \frac{9}{8} - 1 = \frac{1}{8}
\end{aligned}$$

3rd case: Induction step $m - 1 \to m$ for $m \not\equiv 0,\ 1,\ 2 \pmod p$
We use the generic addition formula to compute g_m and h_m:

$$\begin{aligned}
\lambda &= \frac{h_{m-1} - Y}{g_{m-1} - X} \\
g_m &= -g_{m-1} - X + \lambda^2 + (a_1\lambda - a_2) \\
h_m &= -\lambda(g_m - X) - Y - (a_1g_m + a_3)
\end{aligned}$$

Notice that $g_{m-1} - X$ has order -2 in $\mathcal{O}$ because the leading coefficients do not sum to zero:

$$\begin{aligned}
& l(g_{m-1}) - l(X) = 0 \\
\Leftrightarrow\quad & \frac{1}{(m-1)^2} - 1 \equiv 0 \pmod p \quad \text{by the induction hypothesis} \\
\Leftrightarrow\quad & (m-1)^2 \equiv 1 \pmod p \\
\Leftrightarrow\quad & m - 1 \equiv \pm 1 \pmod p \\
\Leftrightarrow\quad & m \equiv 0 \text{ or } 2 \pmod p
\end{aligned}$$

Two cases can be distinguished:

- $(m-1)^3 \not\equiv 1 \pmod{p}$
  Then the leading coefficients of h_{m-1} and Y are different by a similar argument as for g_{m-1} and X, so by the induction hypothesis λ has a simple pole in $\mathcal{O}$. Precisely,

$$l(\lambda) = \frac{l(h_{m-1}) - l(Y)}{l(g_{m-1}) - l(X)} = \frac{\frac{1}{(m-1)^3} - 1}{\frac{1}{(m-1)^2} - 1} = \frac{m^2 - m + 1}{m(m-1)}.$$

  We ignore the term $a_1\lambda - a_2$ of order at least -1 in the equation for g_m; all other terms have order -2, and their leading coefficients do not sum to zero:

$$\begin{aligned} l(g_m) &= -l(g_{m-1}) - l(X) + l(\lambda)^2 \\ &= -\frac{1}{(m-1)^2} - 1 + \left(\frac{m^2 - m + 1}{m(m-1)}\right)^2 = \frac{1}{m^2} \end{aligned}$$

  Similarly for h_m:

$$\begin{aligned} l(h_m) &= -l(\lambda)(l(g_m) - l(X)) - l(Y) \\ &= -\frac{m^2 - m + 1}{m(m-1)} \cdot \frac{1 - m^2}{m^2} - 1 = \frac{1}{m^3} \end{aligned}$$

- $m - 1 \equiv 1 \pmod{p}$
  Then $l(h_{m-1}) = l(Y) = 1$ so that $\mathrm{ord}_{\mathcal{O}}(h_{m-1} - Y) \geq -2$ and λ is regular in $\mathcal{O}$. In the expression for g_m, all terms are regular in $\mathcal{O}$ except for $-g_{m-1} - X$ which is of order -2 and has the leading coefficient

$$-\frac{1}{(m-1)^2} - 1 = -\frac{(m-1) + (m-1)^3}{(m-1)^3} = -m = \frac{-m^3}{m^2}.$$

  Note that from $(m-1)^3 = 1$ it follows that

$$\begin{aligned} 0 &= (m-1)^3 - 1 \\ &= m^3 - 3m^2 + 3m - 2 \\ &= (m^3 + 1)\left(1 - \frac{3}{m+1}\right). \end{aligned}$$

  Since $m + 1 \not\equiv 3 \pmod{p}$ it follows that $-m^3 = 1$ in $\mathbb{F}_p$ and $l(g_m) = \frac{1}{m^2}$. For h_m the leading coefficient is

$$l(h_m) = -l(Y) = -1 = \frac{1}{m^3}.$$

4th case: Induction step $m - 2 \to m$ for $m \equiv 1 \pmod{p}$
Now we can express g_m and h_m by applying the generic addition formula to

$[m-2]$ and $[2]$. By Corollary 3.10 the formula is valid for $m \neq 4$, which is the case for $p \neq 3$.

$$\begin{aligned}
\lambda &= \frac{h_{m-2} - h_2}{g_{m-2} - g_2} \\
g_m &= -g_{m-2} - g_2 + \lambda^2 + (a_1\lambda - a_2) \\
h_m &= -\lambda(g_m - g_2) - h_2 - (a_1 g_m + a_3)
\end{aligned}$$

We proceed exactly as above by proving that λ has a simple pole in $\mathcal{O}$ and considering the terms of order -2 resp. -3 in the expressions for g_m and h_m:

$$\begin{aligned}
l(\lambda) &= \frac{l(h_{m-2}) - l(h_2)}{l(g_{m-2}) - l(g_2)} = \frac{\frac{1}{(m-2)^3} - \frac{1}{8}}{\frac{1}{(m-2)^2} - \frac{1}{4}} \\
&= \frac{\frac{1}{(-1)^3} - \frac{1}{8}}{\frac{1}{(-1)^2} - \frac{1}{4}} = -\frac{3}{2} \neq 0, \infty \\
l(g_m) &= -l(g_{m-2}) - l(g_2) + l(\lambda)^2 \\
&= -\frac{1}{(-1)^2} - \frac{1}{4} + \frac{9}{4} = 1 = \frac{1}{m^2} \\
l(h_m) &= -l(\lambda)(l(g_m) - l(g_2)) - l(h_2) \\
&= \frac{3}{2}\left(1 - \frac{1}{4}\right) - \frac{1}{8} = 1 = \frac{1}{m^3}
\end{aligned}$$

5th case: Induction step $m-3 \to m$ for $m \equiv 2 \pmod p$

In this case we have $m \neq 6$. Since the proposition has been settled for $[3]$ with the third case, we can apply the generic addition formula to $[m-3]$ and $[3]$ to obtain

$$\begin{aligned}
\lambda &= \frac{h_{m-3} - h_3}{g_{m-3} - g_3} \\
g_m &= -g_{m-3} - g_3 + \lambda^2 + (a_1\lambda - a_2) \\
h_m &= -\lambda(g_m - g_3) - h_3 - (a_1 g_m + a_3)
\end{aligned}$$

Now $g_{m-3} - g_3$ is of order -2 since the leading coefficients sum to

$$\frac{1}{(m-3)^2} - \frac{1}{9} = 1 - \frac{1}{9} = \frac{8}{9} \neq 0.$$

The leading coefficients of h_{m-3} and h_3 are -1 and $\frac{1}{27}$, respectively. Again we must distinguish two subcases:

- $p \neq 7$

  Then $l(h_{m-3}) \neq l(h_3)$, and λ has a simple pole in $\mathcal{O}$ and leading coefficient

$$\frac{-1 - \frac{1}{27}}{1 - \frac{1}{9}} = -\frac{7}{6}.$$

Collecting the terms of order -2 in the expression for g_m and summing up the leading coefficients we obtain

$$\begin{aligned} l(g_m) &= -l(g_{m-3}) - l(g_3) + l(\lambda)^2 \\ &= -1 - \frac{1}{9} + \frac{49}{36} = \frac{1}{4} = \frac{1}{m^2} \\ l(h_m) &= -l(\lambda)(l(g_m) - l(g_3)) - l(h_3) \\ &= \frac{7}{6}\left(\frac{1}{4} - \frac{1}{9}\right) - \frac{1}{27} = \frac{1}{8} = \frac{1}{m^3} \end{aligned}$$

- $p = 7$
  Then $l(h_{m-3}) - l(h_3) = -\frac{28}{27} = 0$, and λ is regular in $\mathcal{O}$. We collect the terms of order -2 resp. -3 and sum up their leading coefficients, which yields

$$\begin{aligned} l(g_m) &= -l(g_{m-3}) - l(g_3) \\ &= -1 - \frac{1}{9} = -\frac{10}{9} = \frac{1}{4} = \frac{1}{m^2} \\ l(h_m) &= -l(h_3) = -\frac{1}{27} = \frac{1}{8} = \frac{1}{m^3} \end{aligned}$$

This finishes the proof for $p \neq 2, 3$.

For $p \in \{2, 3\}$, we proceed in steps of p. The case $m = 1$ is trivial, and for $p = 3$ the case $m = 2$ has been shown during the proof of Proposition 3.42. This provides the basis for the induction.

6th case: $p \in \{2, 3\}$, induction step $m - p \to m$ for $m \not\equiv 0 \pmod p$

Let α and γ be as in Proposition 3.42. The addition formulae for $[m - p]$ and $[p]$ are given by

$$\begin{aligned} \lambda &= \frac{h_p - h_{m-p}}{g_p - g_{m-p}} \\ g_m &= -g_{m-p} - g_p + \lambda^2 + a_1\lambda - a_2 \\ h_m &= -\lambda(g_m - g_p) - h_p - (a_1 g_m + a_3) \end{aligned}$$

By Proposition 3.42, λ^2 and g_p both have a pole of order 2α in $\mathcal{O}$. We claim that

$$\begin{aligned} \mathrm{ord}_{\mathcal{O}}(g_m - g_{m-p}) &= \alpha - 3 \text{ for } p = 3 \\ l(g_m - g_{m-p}) &= \frac{\gamma}{m^3} \text{ for } p = 3 \\ \mathrm{ord}_{\mathcal{O}}(g_m - g_{m-p}) &\geq \alpha - 2 \text{ for } p = 2. \end{aligned}$$

Proof of the claim:

$$g_m - g_{m-p} = -2g_{m-p} - g_p + \lambda^2 + a_1\lambda - a_2 = \frac{r}{s}$$

by the addition formulae above, where r and s are given by

$$\begin{aligned}
s &= g_p^2 - 2g_pg_{m-p} + g_{m-p}^2 \\
r &= -(g_p^2 - 2g_pg_{m-p} + g_{m-p}^2)(g_p + 2g_{m-p} + a_2) \\
&\quad +(h_p^2 - 2h_ph_{m-p} + h_{m-p}^2) + a_1(h_p - h_{m-p})(g_p - g_{m-p}) \\
&= (h_p^2 + a_1h_pg_p - g_p^3 - a_2g_p^2) \\
&\quad +(h_{m-p}^2 + a_1h_{m-p}g_{m-p} - 2g_{m-p}^3 - a_2g_{m-p}^2) \\
&\quad +3g_pg_{m-p}^2 + 2a_2g_pg_{m-p} - 2h_ph_{m-p} - a_1h_pg_{m-p} - a_1h_{m-p}g_p \\
&= -a_3h_p + a_4g_p - a_3h_{m-p} - g_{m-p}^3 + a_4g_{m-p} + 2a_6 \\
&\quad +3g_pg_{m-p}^2 + 2a_2g_pg_{m-p} - 2h_ph_{m-p} - a_1h_pg_{m-p} - a_1h_{m-p}g_p
\end{aligned}$$

since (g_p, h_p) and (g_{m-p}, h_{m-p}) satisfy the equation of the elliptic curve. Notice that the orders in $\mathcal{O}$ and leading coefficients of g_p and h_p are known by Proposition 3.42, and those of g_{m-p} and h_{m-p} are known by the induction hypothesis. The denominator s has order -4α and leading coefficient $\frac{1}{\gamma^4}$. For $p = 3$, the unique term of lowest order in the last expression for r is $-2h_ph_{m-p} = h_ph_{m-p}$; its order is $-3\alpha - 3$ and its leading coefficient is $\frac{1}{\gamma^3(m-p)^3} = \frac{1}{\gamma^3m^3}$. For $p = 2$, this term vanishes, and the orders of the remaining terms are at least $-3\alpha - 2$. This proves the claim via Proposition 2.14.

Since $\alpha \geq p$, it follows in particular that $\mathrm{ord}_{\mathcal{O}}(g_m - g_{m-p}) \geq 0$. Hence Proposition 2.14 and the remark after Definition 3.40 imply that $\mathrm{ord}_{\mathcal{O}}\, g_m = \mathrm{ord}_{\mathcal{O}}\, g_{m-p} = -2$ and $l(g_m) = l(g_{m-p}) = \frac{1}{(m-p)^2} = \frac{1}{m^2}$ as desired.

For h_m we collect all terms of order at most -3 in the addition formula to obtain

$$\begin{aligned}
-\lambda(g_m - g_p) - h_p &= \frac{-(h_p - h_{m-p})(g_m - g_p) - h_p(g_p - g_{m-p})}{g_p - g_{m-p}} \\
&= \frac{-h_p(g_m - g_{m-p}) - h_{m-p}g_p + h_{m-p}g_m}{g_p - g_{m-p}}.
\end{aligned}$$

The denominator has order -2α in $\mathcal{O}$ with leading coefficient $\frac{1}{\gamma^2}$. We show that the numerator has order $-2\alpha - 3$ and leading coefficient $\frac{1}{\gamma^2m^3}$, which shows the desired result for h_m. Assume first that $p = 3$. Then in the numerator, we have

$$\begin{aligned}
\mathrm{ord}_{\mathcal{O}}(h_p(g_{m-p} - g_m)) &= -2\alpha - 3 \text{ by the claim} \\
\mathrm{ord}_{\mathcal{O}}(h_{m-p}g_m) &= -5 \\
\mathrm{ord}_{\mathcal{O}}(h_{m-p}g_p) &= -2\alpha - 3.
\end{aligned}$$

The first and the last term have minimal order, and their leading coefficients do not sum to zero. Precisely,

$$l(-h_p(g_m - g_{m-p})) + l(-h_{m-p}g_p) = -\frac{1}{\gamma^3} \cdot \frac{\gamma}{m^3} - \frac{1}{m^3} \cdot \frac{1}{\gamma^2} = \frac{1}{\gamma^2m^3}$$

by the claim. For $p = 2$, the claim shows that the unique term with minimal order in the numerator is $h_{m-p}g_p$ with leading coeffient $-\frac{1}{\gamma^2 m^3} = \frac{1}{\gamma^2 m^3}$.

This finishes the proof of Proposition 3.43. □

In characteristic 2 or 3 we can exploit the preparations made with Proposition 3.42 to state a more general result.

Proposition 3.44 *Let $p \in \{2, 3\}$, $m = p^\nu m'$ a positive integer with p and m' coprime. Then*

$$\begin{aligned} \operatorname{ord}_{\mathcal{O}} g_m &= -2\alpha^\nu \\ \operatorname{ord}_{\mathcal{O}} h_m &= -3\alpha^\nu \\ e_{[m]} &= \alpha^\nu, \end{aligned}$$

where α is defined as in Proposition 3.42.

Proof: Again, we prove the assertions by induction on ν. For $\nu = 0$, they are just Propositions 3.43 and 3.35. Now we assume the assertions to be correct for $m = p^\nu m'$ and proceed inductively to pm:

$$\begin{aligned} \operatorname{ord}_{\mathcal{O}} g_{pm} &= \operatorname{ord}_{\mathcal{O}}(g_p \circ [m]) \\ &= e_{[m]} \operatorname{ord}_{\mathcal{O}} g_p \quad \text{by Lemma 3.14} \\ &= \alpha^\nu(-2\alpha) \quad \text{by the induction hypothesis} \\ &= -2\alpha^{\nu+1} \end{aligned}$$

By the same argument, $\operatorname{ord}_{\mathcal{O}} h_{pm} = -3\alpha^{\nu+1}$. Recall from Theorem 2.26 that $\frac{X}{Y}$ is a uniformising parameter for $\mathcal{O}$. Then

$$e_{[pm]} = \operatorname{ord}_{\mathcal{O}}\left(\frac{X}{Y} \circ [pm]\right) = \operatorname{ord}_{\mathcal{O}}\left(\frac{g_{pm}}{h_{pm}}\right) = \alpha^{\nu+1}.$$

□

It is not difficult to show that in the setting of the previous proposition with $p = 2$ the leading coefficients are as follows:

	$j \neq 0$	j = 0
$l(g_m)$	$\frac{1}{a_1^{2(2^\nu-1)}}$	$\frac{1}{a_3^{\frac{2}{3}(4^\nu-1)}}$
$l(h_m)$	$\frac{1}{a_1^{3(2^\nu-1)}}$	$\frac{1}{a_3^{4^\nu-1}}$

We renounce at a proof since we do not need the result in the remainder of this book.

The next step is to examine the divisor of $g_m - g_n$, where m and n satisfy some extra conditions. By $\langle E[m] \rangle$ we denote the divisor with coefficients 1 at the points of $E[m]$ and zero at all other points.

Proposition 3.45 *Let m and n be non-zero integers.*

- *If $p \neq 2, 3$ and $m, n, m+n$ and $m-n$ are coprime to p, then*

$$\operatorname{div}(g_m - g_n) = \langle E[m+n] \rangle + \langle E[m-n] \rangle - 2\langle E[m] \rangle - 2\langle E[n] \rangle$$

- *If $p \in \{2,3\}$, m is coprime to p and $n = p^\nu n'$ with $\nu \geq 1$ and n' coprime to p, then*

$$\operatorname{div}(g_m - g_n) = \langle E[m+n] \rangle + \langle E[m-n] \rangle - 2\langle E[m] \rangle - 2\alpha^\nu \langle E[n] \rangle,$$

where α is the number of Proposition 3.42.

Proof: To shorten the presentation we give a proof only for the case $p \in \{2,3\}$ and leave the remaining case, which needs very similar arguments, as an exercise to the reader. (It can also be found in [Charlap and Robbins, 1988], Theorem 7.7, p. 39.) First note that since g_m and g_n have poles exactly in the points of $E[m]$ resp. $E[n]$, $g_m - g_n$ can have poles only in $E[m] \cup E[n]$. Concerning the zeros, $(g_m - g_n)(P) = 0$ means $X(mP) = X(nP)$, or equivalently $mP = \pm nP$, so $g_m - g_n$ can have zeros only in $E[m+n] \cup E[m-n]$. Thus it is sufficient to determine the order of $g_m - g_n$ in the m-, n-, $m+n$- and $m-n$-torsion points. Let P be a point of $E[m] \cup E[n] \cup E[m+n] \cup E[m-n]$.

1st case: $P = \mathcal{O}$
$\operatorname{ord}_{\mathcal{O}} g_m = -2$ by Proposition 3.43 and $\operatorname{ord}_{\mathcal{O}} g_n = -2\alpha^\nu < -2$ by Proposition 3.44. Hence $\operatorname{ord}_{\mathcal{O}}(g_m - g_n) = -2\alpha^\nu$.

2nd case: $P \in (E[m] \cap E[n]) \backslash \{\mathcal{O}\}$
Then $P \in E[m+n] \cap E[m-n]$, so we must show that $\operatorname{ord}_P(g_m - g_n) = 1 + 1 - 2 - 2\alpha^\nu = \operatorname{ord}_{\mathcal{O}}(g_m - g_n)$. Since g_m and g_n are invariant under translation by points of $E[m] \cap E[n]$, we have

$$\begin{aligned} \operatorname{ord}_P(g_m - g_n) &= \operatorname{ord}_P((g_m - g_n) \circ \tau_{-P}) \\ &= e_{\tau_{-P}}(P) \operatorname{ord}_{\tau_{-P}(P)}(g_m - g_n) \quad \text{by Lemma 3.14} \\ &= \operatorname{ord}_{\mathcal{O}}(g_m - g_n) \quad \text{by Lemma 3.16} \end{aligned}$$

3rd case: $P \in E[m] \backslash E[n]$
It follows that $P \notin E[m+n] \cup E[m-n]$, and we have to show that $\operatorname{ord}_P(g_m - g_n) = -2$. Since g_m is invariant under translation by P and m is coprime to p, $\operatorname{ord}_P g_m = \operatorname{ord}_{\mathcal{O}} g_m = -2$ by a computation similar to that of the second case; g_n is regular at $P \notin E[n]$, so $\operatorname{ord}_p(g_m - g_n) = -2$.

4th case: $P \in E[n] \backslash E[m]$
We have that

$$\operatorname{ord}_P g_m \geq 0 \quad \text{and} \quad \operatorname{ord}_P g_n = -2\alpha^\nu$$

by Proposition 3.44, so

$$\operatorname{ord}_P(g_m - g_n) = -2\alpha^\nu$$

as desired.

From now on we assume $P \notin E[m] \cup E[n]$.

5^{th} case: $P \in E[m-n] \backslash E[m+n]$, i.e. $mP = nP \neq -nP$
We have to show that $g_m - g_n$ has a simple zero in P. It is obvious that P is a zero of $g_m - g_n$; to determine its multiplicity consider the derivation

$$\begin{aligned} D(g_m - g_n) &= m(2h_m + a_1 g_m + a_3) - n(2h_n + a_1 g_n + a_3) \\ &\quad \text{by Theorem 3.27} \\ &= m(2h_m + a_1 g_m + a_3) \quad \text{since } p|n. \end{aligned}$$

Consequently

$$\begin{aligned} D(g_m - g_n)(P) &= m(2h_m(P) + a_1 g_m(P) + a_3) \\ &= m(2Y + a_1 X + a_3)(mP) \\ &\neq 0; \end{aligned}$$

otherwise $(2Y + a_1 X + a_3)(mP) = 0$ because m and p are coprime, and $mP = nP$ would be of order 2, contradicting $nP \neq -nP$. So $\text{ord}_P(D(g_m - g_n)) = 0$, and $\text{ord}_P(g_m - g_n) = 1$ by Theorem 3.30.

6^{th} case: $P \in E[m+n] \backslash E[m-n]$, i.e. $mP = -nP \neq nP$
Then P is a simple zero of $g_m - g_n$ with an analogous argumentation as in the previous case.

7^{th} case: $P \in E[m+n] \cap E[m-n]$, i.e. $mP = nP$ is of order 2
This is the most difficult case. Again it is easy to see that P is a zero of $g_m - g_n$, and we have to show that its multiplicity is 2. We must consider the cases $p = 3$ and $p = 2$ separately:

- $p = 3$
  By Theorem 3.30, $\text{ord}_P(g_m - g_n) = 2$ is equivalent to $\text{ord}_P(D(g_m - g_n)) = 1$ and $\text{ord}_P(D^2(g_m - g_n)) = 0$.

$$\begin{aligned} D(g_m - g_n) &= m(-h_m + a_1 g_m + a_3) \quad \text{as above, and} \\ D(g_m - g_n)(P) &= 0 \quad \text{since } mP \text{ is of order 2.} \\ D^2(g_m - g_n) &= m^2(a_2 g_m - a_4 + a_1 h_m + a_1(-h_m + a_1 g_m + a_3)) \\ &\quad \text{by Theorem 3.27} \\ D^2(g_m - g_n)(P) &= m^2(a_2 g_m - a_4 + a_1 h_m)(P) \\ &= m^2 \frac{\partial E}{\partial X}(mP) \\ &\neq 0 \end{aligned}$$

  because $\frac{\partial E}{\partial Y}$ is zero in mP and E is non-singular.

- For $p = 2$ this procedure cannot work because by Theorem 3.30 derivation does not reduce the order 2. Instead we examine $a_1 g_m + a_3$ and $a_1 g_n + a_3$:

$$\begin{aligned} \operatorname{ord}_P(a_1 g_m + a_3) &= \operatorname{ord}_P((a_1 X + a_3) \circ [m]) \\ &= e_{[m]} \operatorname{ord}_{mP}(a_1 X + a_3) \quad \text{by Lemma 3.14.} \end{aligned}$$

Note that $e_{[m]} = 1$ by Proposition 3.35 and that mP is a point of order 2, so $\operatorname{ord}_{mP}(a_1 X + a_3) = 2$.

$$\begin{aligned} \operatorname{ord}_P(a_1 g_n + a_3) &= e_{[n]} \operatorname{ord}_{nP}(a_1 X + a_3) \\ &= \alpha^{\nu} \cdot 2 \quad \text{by Proposition 3.44} \\ &> 2 \end{aligned}$$

So $\operatorname{ord}_P(g_m - g_n) = \operatorname{ord}_P((a_1 g_m + a_3) - (a_1 g_n + a_3)) = 2$; note that a_1 cannot be zero in this case because there is a point of order 2.

□

Proposition 3.46 *If an integer m is coprime to p, then*

$$|E[m]| = m^2.$$

Proof: Only during this proof, denote $|E[m']|$ by $d_{m'}$. Suppose first that $p \neq 2, 3$. Recall that principal divisors have degree zero by Theorem 2.28, so taking degrees in Proposition 3.45 it follows that

$$d_{m'+n'} + d_{m'-n'} - 2d_{m'} - 2d_{n'} = 0 \tag{3.19}$$

for p coprime to $m', n', m' + n'$ and $m' - n'$. We proceed by induction on m. The result is clear for $m = 1$ and $m = 2$. Assume now $d_r = r^2$ for $r < m$ and $p \nmid r$.

1st case: $m \not\equiv 1, 2 \pmod p$
Choose $n' = 1$ in (3.19) and replace m' by $m - 1$ to obtain

$$\begin{aligned} d_m &= 2d_{m-1} + 2d_1 - d_{m-2} \\ &= 2(m-1)^2 + 2 - (m-2)^2 = m^2. \end{aligned}$$

2nd case: $m \equiv 1 \pmod p$, $m \geq p + 1$
Since $p \geq 5$ we conclude that $m \not\equiv 2$ and $m \not\equiv 4 \pmod p$; choose $n' = 2$ and replace m' in (3.19) by $m - 2$:

$$\begin{aligned} d_m &= 2d_{m-2} + 2d_2 - d_{m-4} \\ &= 2(m-2)^2 + 8 - (m-4)^2 = m^2. \end{aligned}$$

3rd case: $m \equiv 2 \pmod{p}$, $m \geq p+2$
Then $m \not\equiv 3$ and $m \not\equiv 6 \pmod{p}$. Choose $n' = 3$ and replace m' in (3.19) by $m-3$:

$$\begin{aligned} d_m &= 2d_{m-3} + 2d_3 - d_{m-6} \\ &= 2(m-3)^2 + 18 - (m-6)^2 = m^2. \end{aligned}$$

Note that we know d_3 by the first case.

To prove the assertion for $p \in \{2,3\}$ and m coprime to p we choose $n = p$ and take degrees in Proposition 3.45:

$$d_m = 2d_{m-p} + 2\alpha d_p - d_{m-2p}$$

Note that by definition of α, regardless of the value of j, $\alpha d_p = p^2$, so

$$\begin{aligned} d_m &= 2(m-p)^2 + 2p^2 - (m-2p)^2 \quad \text{by induction} \\ &= m^2. \end{aligned}$$

To start the induction we need the result for $m = -1$ and $m = 1$ in characteristic 2 and $m \in \{-2,-1,1,2\}$ in characteristic 3, where they are well-known. □

We have now all ingredients to prove the Propositions 3.37 and 3.38. To pass from the information on the number of points in a group — which we have for $E[m]$ with m coprime to p — to the group structure we use the following theorem, whose proof can be found in any algebra text book.

Theorem 3.47 (Fundamental Theorem on Abelian Groups) *Let $G \neq \{0\}$ be an abelian group. Then there are unique positive integers r and $n_1, \ldots, n_r \geq 2$ such that $n_i | n_{i+1}$ for $i = 1, \ldots, r-1$ and*

$$G \simeq \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_r}.$$

Proof of Proposition 3.37: We use induction over the number of prime factors of m and apply that $|E[m]| = m^2$.

1st case: $m = q$ is prime
Then by the Fundamental Theorem $E[q] \simeq \mathbb{Z}_q \times \mathbb{Z}_q$, or $E[q] \simeq \mathbb{Z}_{q^2}$ is cyclic. But in the second case, $E[q]$ contains a point of order q^2, which is not annulled by multiplication with q, a contradiction.

2nd case: $m = qm'$ with q a prime, $|m'| > 1$
Write $E[m] \simeq \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_r}$ as in the Fundamental Theorem. From $E[m] = \{P : m'qP = \mathcal{O}\}$ and the surjectivity of $[q]$ (see Proposition 3.3) we conclude that

$$\begin{aligned} E[m'] &= \{P : m'P = \mathcal{O}\} \\ &= \{qP : m'qP = \mathcal{O}\} \\ &= \{qP : P \in E[m]\} \\ &= qE[m]. \end{aligned}$$

Hence

$$\begin{aligned} E[m'] &\simeq q\mathbb{Z}_{n_1} \times \cdots \times q\mathbb{Z}_{n_r} \\ &\simeq \mathbb{Z}_{b_1} \times \cdots \times \mathbb{Z}_{b_r} \end{aligned}$$

with

$$b_i = \begin{cases} n_i & \text{if } q \nmid n_i \\ \frac{n_i}{q} & \text{if } q | n_i \end{cases}$$

and $b_i | b_{i+1}$ since $n_i | n_{i+1}$. Applying the induction hypothesis to m' we see from the uniqueness of the b_i that $b_1 = \cdots = b_{r-2} = 1$ and $b_{r-1} = b_r = m'$, so $n_1 = \cdots = n_{r-2} = q$ and $n_{r-1} = n_r = qm' = m$. It follows that $m^2 = |E[m]| = q^{r-2}m^2$, which proves that $r = 2$.

□

To prove Proposition 3.38 we need the following lemma.

Lemma 3.48 *The number of p-torsion points is less than p^2.*

Proof: We know the result for $p \in \{2, 3\}$, and for arbitrary characteristic it will be proved in Section 3.6 about division polynomials (see page 89). □

Proof of Proposition 3.38: As an immediate consequence of the lemma we have $E[p] = \{O\}$ or $E[p] \simeq \mathbb{Z}_p$. We assume $E[p] \simeq \mathbb{Z}_p$ and proceed by induction on ν, assuming that $E[p^{\nu-1}] \simeq \mathbb{Z}_{p^{\nu-1}}$. Consider the group homomorphism

$$\rho : E[p^\nu] \to E[p^{\nu-1}], \quad P \mapsto pP,$$

which is the restriction of $[p]$ to $E[p^\nu]$. Since $[p]$ is surjective and the preimage of a $p^{\nu-1}$-torsion point under $[p]$ must be of p^ν-torsion, ρ is surjective. Its kernel is $E[p]$, so $|E[p^\nu]| = |\operatorname{im}\rho|\,|\ker\rho| = p^{\nu-1}p = p^\nu$. Let Q be a point of order $p^{\nu-1}$ which exists by the induction hypothesis. Then any of its preimages under ρ is of order p^ν, so $E[p^\nu]$ must be cyclic. □

The two propositions can be combined to yield Theorem 3.39 via the following lemma from elementary group theory:

Lemma 3.49 *If m and n are coprime, then $E[mn] \simeq E[m] \times E[n]$.*

Proof: The proof is the exact analogue of the Chinese Remainder Theorem and constructs the isomorphism explicitly. By the Euclidian algorithm compute integers e and f such that $em + fn = 1$. Define group homomorphisms

$$\iota : E[m] \times E[n] \to E[mn], \quad (P, Q) \mapsto P + Q$$

$$\pi : E[mn] \to E[m] \times E[n], \quad P \mapsto (fnP, emP)$$

It is easy to see that ι and π are well-defined and satisfy the homomorphism property. Moreover $\pi \circ \iota = \mathrm{id}\,|_{E[m]\times E[n]}$ and $\iota \circ \pi = \mathrm{id}\,|_{E[mn]}$:

$$\begin{aligned}
\pi \circ \iota(P,Q) &= (fn(P+Q), em(P+Q)) \\
&= (fnP, emQ) \quad \text{since } Q \in E[n],\ P \in E[m] \\
&= ((fn+em)P, (fn+em)Q) \quad \text{for the same reason} \\
&= (P,Q) \\
\iota \circ \pi(P) &= (fn+em)P = P
\end{aligned}$$

□

Proof of Theorem 3.39: Using the lemma, we have for $m = p^\nu m'$, $p \nmid m'$, that

$$E[m] \simeq E[m'] \times E[p^\nu].$$

If $E[p] = \{\mathcal{O}\}$, then $E[p^\nu] = \{\mathcal{O}\}$, and

$$E[m] \simeq E[m'] \simeq \mathbb{Z}_{m'} \times \mathbb{Z}_{m'}.$$

Otherwise

$$E[m] \simeq \mathbb{Z}_{m'} \times \mathbb{Z}_{m'} \times \mathbb{Z}_{p^\nu} \simeq \mathbb{Z}_{m'} \times \mathbb{Z}_m$$

by the Chinese Remainder Theorem. □

3.6 DIVISION POLYNOMIALS

To obtain information on the m-torsion points we have so far examined the rational functions g_m and h_m, which have poles exactly at these points. However, we have no information on their zeros, and the poles are double at least for m coprime to p. This section aims at reducing this "noise" by defining rational functions with simple zeros exactly at the m-torsion points and a pole only at $\mathcal{O}$. If such a function exists it must be a polynomial by Proposition 2.34. By the isomorphism of E with the degree zero part of the Picard group (Corollary 2.47) such a polynomial exists if the m-torsion points sum to $\mathcal{O}$. This is indeed the case for m and p coprime: To any m-torsion point P, its inverse $\overline{P}$ is an m-torsion point, too. So the sum over all m-torsion points which are not of order 2 is $\mathcal{O}$. If $E[m]$ contains a point of order 2, then m must be even, $E[2] \subseteq E[m]$ and $p \neq 2$. In this case there are three points of order 2, which sum to $\mathcal{O}$ since there is a rational function with divisor $\langle E[2]\rangle - 4\langle\mathcal{O}\rangle$, namely the line $2Y + a_1X + a_3$.

To avoid complications with positive characteristic we define these polynomials first in characteristic zero and verify that the essential properties still hold in characteristic $p > 0$ after reducing modulo p.

SO UNTIL FURTHER NOTICE WE ASSUME $p = 0$.

Definition 3.50 *For an integer $m \neq 0$ the m-*th division polynomial *ψ_m is the unique rational function with divisor $\langle E[m]\rangle - m^2\langle \mathcal{O}\rangle$ and leading coefficient m. By convention $\psi_0 = 0$.*

Proposition 3.51 *The following identities are valid for positive integers m and n:*

1. $\psi_{-m} = -\psi_m$
2. $\psi_m^2 = m^2 \sum_{P \in E[m]\setminus\{\mathcal{O}\}} (X - X(P))$
3. $\psi_m \in \begin{cases} K[X] & \textit{if } m \textit{ is odd} \\ (2Y + a_1X + a_3)K[X] & \textit{if } m \textit{ is even} \end{cases}$
4. $\psi_m\psi_n \in K[X]$ *if m and n have the same parity.*

Proof:

1. This is immediately clear from $E[m] = E[-m]$ and Definition 3.50.
2. Recall that $\operatorname{div}(X - X(P)) = \langle P\rangle + \langle \overline{P}\rangle - 2\langle \mathcal{O}\rangle$. It follows that the divisors of the two rational functions agree. Moreover, both leading coefficients are m^2, so the rational functions must be the same by Corollary 2.35.
3. If m is odd, then $E[m]$ contains no point of order 2. Decompose $E[m] = S \dot{\cup} \overline{S} \dot{\cup} \{\mathcal{O}\}$ where $\overline{S} = \{\overline{P} : P \in S\}$. Then $\psi_m = m \sum_{P\in S} (X - X(P))$ with the same argument as in 2. If m is even, then $E[2] \subseteq E[m]$. Decompose $E[m] = S \dot{\cup} \overline{S} \dot{\cup} E[2]$ such that $\psi_m = \frac{m}{2}\psi_2 \sum_{P\in S} (X - X(P))$. In the introduction to this section we have already seen that $\psi_2 = 2Y + a_1X + a_3$, which finishes the proof of this part.
4. By 3. it is sufficient to see that $\psi_2^2 \in K[X]$.

$$\begin{aligned} \psi_2^2 &= (2Y + a_1X + a_3)^2 \\ &= 4Y^2 + 4Y(a_1X + a_3) + (a_1X + a_3)^2 \\ &= 4(X^3 + a_2X^2 + a_4X + a_6) + (a_1X + a_3)^2 \\ &\in K[X] \end{aligned}$$

□

Proposition 3.52

$$g_m - g_n = -\frac{\psi_{m+n}\psi_{m-n}}{\psi_m^2\psi_n^2} \quad \textit{for } m, n \neq 0$$

Proof: By Proposition 3.45 both sides have the same divisor. Their leading coefficients are $\frac{1}{m^2} - \frac{1}{n^2}$ by Proposition 3.43 and $-\frac{(m+n)(m-n)}{m^2n^2}$ by definition, so they are the same. Hence the two rational functions agree. □

Proposition 3.53 *The division polynomials are the unique polynomials which satisfy the following recursion:*

$$\begin{aligned}\psi_0 &= 0\\ \psi_1 &= 1\\ \psi_2 &= 2Y + a_1X + a_3\\ \psi_3 &= 3X^4 + b_2X^3 + 3b_4X^2 + 3b_6X + b_8\\ \frac{\psi_4}{\psi_2} &= 2X^6 + b_2X^5 + 5b_4X^4 + 10b_6X^3 + 10b_8X^2\\ &\quad +(b_2b_8 - b_4b_6)X + (b_4b_8 - b_6^2),\end{aligned}$$

where the b_i are as in Definition 2.7,

$$\psi_{m+n}\psi_{m-n} = \psi_n^2\psi_{m+1}\psi_{m-1} - \psi_m^2\psi_{n+1}\psi_{n-1} \tag{3.20}$$

Furthermore

$$\begin{aligned}\psi_{2m} &= \frac{\psi_m}{\psi_2}(\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2) & (3.21)\\ &= (\psi_2 \circ [m])\psi_m^4 & (3.22)\\ \psi_{2m+1} &= \psi_{m+2}\psi_m^3 - \psi_{m+1}^3\psi_{m-1} & (3.23)\end{aligned}$$

Proof: The validity of (3.20) for $m, n \neq 0$ follows from Proposition 3.52: Write

$$g_m - g_n = (g_m - g_1) - (g_n - g_1)$$

or equivalently

$$\frac{\psi_{m+n}\psi_{m-n}}{\psi_m^2\psi_n^2} = \frac{\psi_{m+1}\psi_{m-1}}{\psi_m^2} - \frac{\psi_{n+1}\psi_{n-1}}{\psi_n^2}$$

and multiply this equation by $\psi_m^2\psi_n^2$. For $m = 0$ or $n = 0$, (3.20) is trivial. The formulae for ψ_0, ψ_1 and ψ_2 are clear. For ψ_3 we apply Proposition 3.52 with $m = 2$ and $n = 1$, which yields

$$\psi_3 = -(g_2 - X)\psi_2^2,$$

and substitute the known formula for g_2. In a similar way ψ_4 is computed from

$$\psi_4 = -\frac{(g_3 - X)\psi_3^2}{\psi_2}.$$

The uniqueness is verified by evaluating (3.20) with specific values for m and n. Letting $n = 0$ we find that $-\psi_m^2\psi_{-1} = \psi_m^2$ or $\psi_{-1} = -1$. Then choosing $m = 0$ yields $-\psi_n^2 = \psi_n\psi_{-n}$ or $\psi_{-n} = -\psi_n$. Finally for $m \geq 3$ we can fix $n = 2$ and compute ψ_{m+2} inductively as

$$\psi_{m+2} = \frac{\psi_2^2\psi_{m+1}\psi_{m-1} - \psi_m^2\psi_3}{\psi_{m-2}}.$$

So (3.20) together with the fixed values of ψ_0, ψ_1, ψ_2, ψ_3 and ψ_4 determines the division polynomials uniquely.

Equation (3.21) follows from (3.20) by replacing m by $m+1$ and n by $m-1$, (3.23) follows from (3.20) replacing n by m and m by $m+1$. For (3.22) we compare divisors:

$$\begin{aligned}
\operatorname{div}\psi_{2m} &= \langle E[2m]\rangle - 4m^2\langle \mathcal{O}\rangle \\
\operatorname{div}\psi_m^4 &= 4\langle E[m]\rangle - 4m^2\langle \mathcal{O}\rangle \\
\operatorname{div}(\psi_2 \circ [m]) &= \operatorname{div}([m]^*(\psi_2)) \\
&= [m]^*(\operatorname{div}\psi_2) \quad \text{by Proposition 3.13} \\
&= [m]^*(\langle E[2]\rangle - 4\langle \mathcal{O}\rangle) \\
&= \sum_{P\in[m]^{-1}(E[2])} \langle P\rangle - 4\langle E[m]\rangle \\
&= \langle E[2m]\rangle - 4\langle E[m]\rangle,
\end{aligned}$$

so the divisors of ψ_{2m} and $(\psi_2 \circ [m])\psi_m^4$ are the same. We compare the leading coefficients:

$$\begin{aligned}
l(\psi_{2m}) &= 2m \\
l((\psi_2 \circ [m])\psi_m^4) &= l(2h_m + a_1 g_m + a_3)m^4 \\
&= \frac{2}{m^3}m^4 \quad \text{by Proposition 3.43} \\
&= l(\psi_{2m})
\end{aligned}$$

Hence the two sides of the equation represent the same rational function. □

Corollary 3.54

$$\psi_m \in \mathbb{Z}[X,Y,a_1,a_3,a_2,a_4,a_6]/(E) \quad \textit{for all } m$$

More precisely,

$$\psi_m \in \mathbb{Z}[X,a_1,a_3,a_2,a_4,a_6]/(E) \quad \textit{for } m \textit{ odd}$$

and

$$\frac{\psi_m}{\psi_2} \in \mathbb{Z}[X,a_1,a_3,a_2,a_4,a_6]/(E) \quad \textit{for } m \textit{ even.}$$

Proof: We preliminarily verify that $\psi_2^2 \in \mathbb{Z}[X,a_1,a_3,a_2,a_4,a_6]/(E)$ (see the proof of Proposition 3.51, 4.). Then the corollary is proved by induction on m. It is trivial for $m=1$ and $m=3$. For odd $m \geq 5$ it follows from (3.23). For $m=2k$ we use (3.21) and consider the cases k even and k odd separately. □

There is one loose end to fix, namely how to compute h_m in terms of the division polynomials, if this is possible. We can use (3.21) and (3.22):

$$\begin{aligned}
2h_m + a_1 g_m + a_3 &= \psi_2 \circ [m] \\
&= \frac{\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2}{\psi_2\psi_m^3}
\end{aligned}$$

and solve for h_m, but this does not generalise to characteristic 2. In even characteristic we need a different approach, suggested in [Koblitz, 1991]; the result is valid in any characteristic.

Proposition 3.55 *Let m be a positive integer. Then h_m can be expressed in two ways:*

1.

$$h_m = \frac{\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2}{2\psi_2\psi_m^3} - \frac{1}{2}(a_1 g_m + a_3)$$

2.

$$h_m - Y = \frac{\psi_{m-2}\psi_{m+1}^2}{\psi_2\psi_m^3} + (3X^2 + 2a_2X + a_4 - a_1Y)\frac{\psi_{m-1}\psi_{m+1}}{\psi_2\psi_m^2} + \psi_2 \circ [m],$$

or equivalently

$$h_m - Y = \frac{\psi_{m-1}^2\psi_{m+2}}{\psi_2\psi_m^3} + (3X^2 + 2a_2X + a_4 - a_1Y)\frac{\psi_{m-1}\psi_{m+1}}{\psi_2\psi_m^2}$$

Proof: 1. has already been shown above. We abridge $3X^2 + 2a_2X + a_4 - a_1Y$ by s' and prove the first formula of 2. inductively. It is clear for $m = 1$ and can be verified for $m = 2$ by a simple computation with a symbolic algebra programme. For $m > 2$ we express as usual h_m by means of the addition formula:

$$h_m = -\frac{h_{m-1} - Y}{g_{m-1} - X}(g_m - X) - (a_1 g_m + a_3 + Y),$$

so

$$h_m - Y = -\frac{g_m - X}{g_{m-1} - X}(h_{m-1} - Y) - \psi_2 \circ [m] + 2(h_m - Y)$$
$$\Leftrightarrow \quad h_m - Y = \frac{g_m - X}{g_{m-1} - X}(h_{m-1} - Y) + \psi_2 \circ [m].$$

We know $h_{m-1} - Y$ by the induction hypothesis and $g_{m-1} - X$ and $g_m - X$ by Proposition 3.52. Combining the results we obtain

$$\frac{g_m - X}{g_{m-1} - X}(h_{m-1} - Y) = \frac{\psi_{m+1}\psi_{m-1}^3}{\psi_{m-2}\psi_m^3}\left(\frac{\psi_m^2\psi_{m-3}}{\psi_2\psi_{m-1}^3} + s'\frac{\psi_m\psi_{m-2}}{\psi_2\psi_{m-1}^2} + \psi_2 \circ [m-1]\right)$$

$\psi_2 \circ [m-1]$ can be transformed by Equations (3.21) and (3.22) into

$$\psi_2 \circ [m-1] = \frac{\psi_{m+1}\psi_{m-2}^2 - \psi_{m-3}\psi_m^2}{\psi_2\psi_{m-1}^3}$$

Finally,

$$\begin{aligned}\frac{g_m - X}{g_{m-1} - X}(h_{m-1} - Y) &= \frac{\psi_{m+1}\psi_{m-3}}{\psi_2\psi_{m-2}\psi_m} + s'\frac{\psi_{m+1}\psi_{m-1}}{\psi_2\psi_m^2} \\ &\quad + \frac{\psi_{m+1}^2\psi_{m-2}}{\psi_2\psi_m^3} - \frac{\psi_{m+1}\psi_{m-3}}{\psi_2\psi_{m-2}\psi_m} \\ &= \frac{\psi_{m+1}^2\psi_{m-2}}{\psi_2\psi_m^3} + s'\frac{\psi_{m+1}\psi_{m-1}}{\psi_2\psi_m^2}\end{aligned}$$

as desired.

To verify the second expression of 2., we substitute $\psi_2 \circ [m]$ by Formulae (3.21) and (3.22). □

FROM NOW ON, LET p AGAIN BE ARBITRARY.

Since we have seen in Corollary 3.54 that the division polynomials in characteristic zero have coefficients in $\mathbb{Z}$, we define them in characteristic p by reducing modulo p. Then all equations we have found in characteristic zero stay valid since by Corollary 3.54 they are equations in the quotient field of the integral domain $\mathbb{Z}[X, Y, a_1, a_3, a_2, a_4, a_6]/(E)$ — provided that no denominator is zero after reducing modulo p. Reconsidering our formulae we find that we have to prove the following lemma.

Lemma 3.56 $\psi_m \neq 0$ *for* $m \neq 0$.

Proof: We proceed as usual by induction on $m > 0$. The lemma is clear for $\psi_1 = 1$ and $\psi_2 = 2Y + a_1X + a_3$. (The latter is not zero even in characteristic 2 because then $a_1X + a_3$ is not zero.) Otherwise use Proposition 3.52 to write

$$\psi_m\psi_{m-2} = (X - g_{m-1})\psi_{m-1}^2$$

and note that $\psi_{m-2}, \psi_{m-1} \neq 0$ by the induction hypothesis and $X \neq g_{m-1}$ since $m > 2$. So $\psi_m \neq 0$. □

The remaining point to show is that the division polynomials "keep" their divisor even in positive characteristic — otherwise they would not be of much interest. Indeed this is the case with the usual restriction on m.

Proposition 3.57 *If m is coprime to p, then*

$$\operatorname{div}\psi_m = \langle E[m]\rangle - m^2\langle O\rangle.$$

Proof: In characteristic zero, ψ_m has a pole of order $m^2 - 1$ in O and leading coefficient m. Since m and p are coprime, the leading coefficient m does not vanish when reducing modulo p, so the pole in O stays of order $m^2 - 1$. Hence ψ_m has $m^2 - 1$ zeros, counting multiplicities. Consider again the equation

$$g_m - X = -\frac{\psi_{m+1}\psi_{m-1}}{\psi_m^2}.$$

Since $g_m - X$ has poles exactly in the points of $E[m]$ and ψ_{m+1} and ψ_{m-1} have no finite poles, ψ_m must have finite zeros exactly in the points of $E[m]\backslash\{\mathcal{O}\}$. There are $m^2 - 1$ such points, hence the zeros must be simple, which proves the assertion. □

Proof of Lemma 3.48: Since the leading coefficient of ψ_p vanishes in characteristic p, we conclude that ψ_p has less than $p^2 - 1$ finite zeros. But by the same argument as above, all elements of $E[p]\backslash\{\mathcal{O}\}$ are zeros of ψ_p, so $|E[p]| < p^2$. □

Example. Let $p \in \{2, 3\}$, m coprime to p, $n = p^\nu n'$ with $\nu \geq 1$ and n' coprime to p. Then we know by Proposition 3.45 that

$$\operatorname{div}(g_m - g_n) = \langle E[m+n]\rangle + \langle E[m-n]\rangle - 2\langle E[m]\rangle - 2\alpha^\nu\langle E[n]\rangle.$$

On the other hand

$$g_m - g_n = -\frac{\psi_{m+n}\psi_{m-n}}{\psi_m^2\psi_n^2}$$

where m, $m+n$ and $m-n$ are coprime to p. It follows that

$$\operatorname{div}\psi_n = \alpha^\nu\langle E[n]\rangle - \alpha^\nu|E[n]|\langle\mathcal{O}\rangle.$$

So ψ_n has zeros of multiplicity α^ν at the finite points of $E[n]$.

3.7 THE WEIL PAIRING

We are interested in results on the number of points on an elliptic curve with coordinates in a finite field, which is a global information on the curve; but with our examination of m-torsion points we have obtained only "local" information. The task of passing from "local" to "global" is performed by the Weil pairing, a "bilinear" form on the m-torsion points, in a sense that will be made precise in Section 3.8.

THROUGHOUT THIS SECTION WE ASSUME THAT m IS A POSITIVE INTEGER COPRIME TO p.

Before defining the Weil pairing and stating some of its properties we prove a lemma which implicitly relates properties of the endomorphism $[m]$ and of the field extension $K(E)/[m]^*(K(E))$.

Lemma 3.58 *Let r be a rational function which is invariant under translation by points of $E[m]$. Then there is a rational function s such that $r = s \circ [m]$, which means that $r \in [m]^*(K(E))$.*

Proof: It is not difficult to prove the lemma for any separable endomorphism α (replacing $E[m]$ by $\ker\alpha$), making use of the fact that $K(E)/\alpha^*(K(E))$ is separable. Since we have not proved the latter result we need a different argument. Indeed we show that

$$[K(E) : [m]^*(K(E))] \leq m^2 = \deg[m] \tag{3.24}$$

Inequality (3.24) proves the lemma: Write

$$\begin{aligned} J &= [m]^*(K(E)) = \{t \circ [m] : t \in K(E)\} \\ H &= \{r \in K(E) : r \circ \tau_S = r \, \forall S \in E[m]\}, \end{aligned}$$

such that $J \subseteq H \subseteq K(E)$. H is the fixed field of a group of m^2 automorphisms of $K(E)$, namely the translations by m-torsion points. So $[K(E) : H] = m^2$ by Galois theory; by (3.24), $m^2 = [K(E) : H] \geq [K(E) : J]$, so $H = J$, which is the desired result. (Another result from Galois theory, which we do not need here, is that $K(E)/J$ is indeed separable.)

To prove (3.24) we consider the four fields $K(g_m)$, $K(g_m, h_m)$, $K(X)$ and $K(E)$. Notice that $J = K(X \circ [m], Y \circ [m]) = K(g_m, h_m)$.

Recall the equation from Proposition 3.52 for $n = 1$:

$$g_m - X = -\frac{\psi_{m-1}\psi_{m+1}}{\psi_m^2} \in K(X),$$

since by Proposition 3.51, 4., ψ_m^2 and $\psi_{m-1}\psi_{m+1}$ are elements of $K[X]$. Hence $g_m \in K(X)$ and $K(g_m) \subseteq K(X)$ (which reflects that the X-coordinate of mP is the same as that of $m\overline{P} = \overline{mP}$ and thus independent of the Y-coordinate of P). Next we show the following relations on the degrees of the field extensions:

$$\begin{tikzcd}
 & K(E) \arrow[dl, dash, "2"'] \arrow[dr, dash] & \\
K(X) \arrow[dr, dash, "\leq m^2"'] & & J = K(g_m, h_m) \arrow[dl, dash, "\geq 2"] \\
 & K(g_m) &
\end{tikzcd}$$

That $[K(E) : K(X)] = 2$, is well-known. Concerning $[K(X) : K(g_m)]$, notice that X satisfies the following polynomial in $K(g_m)[T]$, which is again derived from Proposition 3.52:

$$f(T) := T\psi_m^2(T) - (\psi_{m-1}\psi_{m+1})(T) - g_m\psi_m^2(T)$$

In characteristic zero ψ_m^2 has a pole of order $2(m^2 - 1)$ in $\mathcal{O}$, so by Lemma 2.31 its degree is $m^2 - 1$; $\psi_{m-1}\psi_{m+1}$ has a pole of order $(m-1)^2 + (m+1)^2 - 2 = 2m^2$ in $\mathcal{O}$, so its degree is m^2, and $\deg f \leq m^2$. In positive characteristic the degrees of the division polynomials can only decrease such that $\deg f \leq m^2$ still holds, and $[K(X) : K(g_m)] \leq m^2$. Finally, $h_m \notin K(g_m)$. Otherwise, we would have

$h_m \in K(X)$, which in turn would imply $Y(mP) = h_m(P) = h_m(\overline{P}) = Y(\overline{mP})$ for all $P \in E$, which is wrong for mP not of order 2. Thus we have shown that $[K(g_m, h_m) : K(g_m)] \geq 2$. Then

$$[K(E) : J] = \frac{[K(E) : K(X)][K(X) : K(g_m)]}{[J : K(g_m)]} \leq \frac{2m^2}{2}.$$

□

We now proceed to the definition of the Weil pairing; the rest of this section is identical to [Charlap and Robbins, 1988], Section 12, pp. 61–68.

For an m-torsion point T define a rational function g_T such that $\operatorname{div} g_T = [m]^*(\langle T \rangle - \langle \mathcal{O} \rangle)$. This is possible by Corollary 2.47 since $[m]^*(\langle T \rangle - \langle \mathcal{O} \rangle)$ is of degree zero, and the points which form the divisor sum to $\mathcal{O}$: Let T_0 be a preimage of T under $[m]$, which exists by Proposition 3.3. Then

$$\begin{aligned} [m]^*(\langle T \rangle - \langle \mathcal{O} \rangle) &= \sum_{T' \in [m]^{-1}(T)} \langle T' \rangle - \sum_{R \in \ker[m]} \langle R \rangle \\ &= \sum_{R \in E[m]} (\langle T_0 + R \rangle - \langle R \rangle). \end{aligned}$$

Now

$$\sum_{R \in E[m]} (T_0 + R - R) = m^2 T_0 = mT = \mathcal{O}.$$

Definition 3.59 *The* Weil pairing *of m-torsion points is the function*

$$e_m : E[m] \times E[m] \to \mu, \quad (S, T) \mapsto \frac{g_T \circ \tau_S}{g_T},$$

where μ denotes the set of m-th roots of unity in K.

Since g_T is unique up to multiplication by a non-zero constant, $e_m(S, T)$ does not depend on the choice of g_T. Note that

$$\begin{aligned} \operatorname{div}(g_T \circ \tau_S) &= \tau_S^*(\operatorname{div} g_T) \quad \text{by Proposition 3.13} \\ &= \tau_S^*\left(\sum_{R \in E[m]} (\langle T_0 + R \rangle - \langle R \rangle)\right) \\ &= \sum_{R \in E[m]} (\langle T_0 + R - S \rangle - \langle R - S \rangle) \\ &= \operatorname{div} g_T, \end{aligned}$$

so that $e_m(S, T) \in K$. That $e_m(S, T)^m = 1$ will be proved at the end of this section.

Proposition 3.60 *The Weil pairing has the following properties:*

1. *"Bilinearity"*

$$\begin{aligned} e_m(S_1+S_2,T) &= e_m(S_1,T)e_m(S_2,T) \\ e_m(S,T_1+T_2) &= e_m(S,T_1)e_m(S,T_2) \quad \forall S,S_1,S_2,T,T_1,T_2 \in E[m] \end{aligned}$$

2. *Identity*

$$e_m(S,S)=1 \quad \forall S \in E[m]$$

3. *Alternation*

$$e_m(S,T)=e_m(T,S)^{-1} \quad \forall S,T \in E[m]$$

4. *Non-degeneracy*

$$\begin{aligned} e_m(S,T)=1\ \forall S \in E[m] &\Leftrightarrow \quad T=\mathcal{O} \\ e_m(S,T)=1\ \forall T \in E[m] &\Leftrightarrow \quad S=\mathcal{O} \end{aligned}$$

5. *Compatibility with endomorphisms*

$$e_m(\alpha(S),\alpha(T))=e_m(S,T)^{\deg\alpha} \quad \textit{where } \alpha \textit{ is a non-zero endomorphism}$$

Proof:

1. The first assertion can be verified in a straightforward way.

$$\begin{aligned} e_m(S_1+S_2,T) &= \frac{g_T \circ \tau_{S_1+S_2}}{g_T} \\ &= \frac{g_T \circ \tau_{S_1} \circ \tau_{S_2}}{g_T} \\ &= \left(\frac{g_T \circ \tau_{S_1}}{g_T} \circ \tau_{S_2}\right)\frac{g_T \circ \tau_{S_2}}{g_T} \\ &= (e_m(S_1,T)\circ\tau_{S_2})\,e_m(S_2,T) \\ &= e_m(S_1,T)\,e_m(S_2,T) \\ &\ \text{since } e_m(S_1,T) \text{ is a constant} \end{aligned}$$

The second part is more complicated because we must relate $g_{T_1+T_2}$ with g_{T_1} and g_{T_2}. To this purpose consider

$$\begin{aligned} \operatorname{div}\frac{g_{T_1+T_2}}{g_{T_1}g_{T_2}} &= [m]^*(\langle T_1+T_2\rangle - \langle T_1\rangle - \langle T_2\rangle + \langle\mathcal{O}\rangle) \\ &= [m]^*(\operatorname{div} h) \\ &= \operatorname{div}(h\circ[m]) \quad \text{by Proposition 3.13} \end{aligned}$$

for a suitable rational function h because $\langle T_1 + T_2\rangle - \langle T_1\rangle - \langle T_2\rangle + \langle O\rangle$ is principal. Thus

$$\frac{g_{T_1+T_2}}{g_{T_1}g_{T_2}} = c\,h\circ[m] \quad \text{with } c \in K^\times$$

is invariant under translation by points of $E[m]$, and

$$\begin{aligned}
e_m(S, T_1 + T_2) &= \left(\frac{g_{T_1+T_2}}{g_{T_1}g_{T_2}}\circ\tau_S\right)\frac{(g_{T_1}\circ\tau_S)(g_{T_2}\circ\tau_S)}{g_{T_1+T_2}} \\
&= \frac{g_{T_1}\circ\tau_S}{g_{T_1}}\cdot\frac{g_{T_2}\circ\tau_S}{g_{T_2}} \\
&= e_m(S,T_1)e_m(S,T_2)
\end{aligned}$$

2. Let S_0 be a point on E with $mS_0 = S$. Consider $g_S\circ\tau_{iS_0}$ and

$$G := \prod_{i=0}^{m-1}(g_S\circ\tau_{iS_0}).$$

$$\begin{aligned}
\operatorname{div}(g_S\circ\tau_{iS_0}) &= \tau_{iS_0}^*(\operatorname{div} g_S) \quad \text{by Proposition 3.13} \\
&= (\tau_{iS_0}^*\circ[m]^*)(\langle S\rangle - \langle O\rangle) \\
&= ([m]\circ\tau_{iS_0})^*(\langle S\rangle - \langle O\rangle) \quad \text{by Proposition 3.15} \\
&= (\tau_{iS}\circ[m])^*(\langle S\rangle - \langle O\rangle) \\
&= [m]^*(\langle S - iS\rangle - \langle -iS\rangle) \\
\operatorname{div} G &= [m]^*\left(\sum_{i=0}^{m-1}(\langle(1-i)S\rangle - \langle(0-i)S\rangle)\right) \\
&= [m]^*\left(\sum_{i=2-m}^{1}\langle iS\rangle - \sum_{i=1-m}^{0}\langle iS\rangle\right) \\
&= [m]^*(\langle S\rangle - \langle S - mS\rangle) \\
&= 0 \quad \text{since } mS = O
\end{aligned}$$

So G is a constant by Proposition 2.34 and Corollary 2.33, and

$$\begin{aligned}
1 &= \frac{G\circ\tau_{S_0}}{G} \\
&= \frac{g_S\circ\tau_{mS_0}}{g_S\circ\tau_{0S_0}} \\
&= \frac{g_S\circ\tau_S}{g_S} \\
&= e_m(S,S)
\end{aligned}$$

3. This assertion is again verified by straightforward formulae manipulations.

$$\begin{aligned}
1 &= e_m(S+T, S+T) \quad \text{by 2.} \\
&= e_m(S,S)e_m(S,T)e_m(T,S)e_m(T,T) \quad \text{by 1.} \\
&= e_m(S,T)e_m(T,S) \quad \text{by 2.}
\end{aligned}$$

4. By the alternation property it is sufficient to prove the first equivalence. If $T = \mathcal{O}$, then g_T is constant and thus invariant under translation, and $e_m(S,T) = 1$. Assume now that $e_m(S,T) = 1$ for all $S \in E[m]$. This means that g_T is invariant under translation by m-torsion points, hence $g_T = r \circ [m]$ for a suitable rational function r by Lemma 3.58. Now
$$\begin{aligned}[m]^*(\langle T\rangle - \langle \mathcal{O}\rangle) &= \operatorname{div} g_T \\ &= \operatorname{div}(r \circ [m]) \\ &= [m]^* \operatorname{div} r \quad \text{by Proposition 3.13,}\end{aligned}$$
so by the injectivity of $[m]^*$, $\operatorname{div} r = \langle T\rangle - \langle \mathcal{O}\rangle$. Since r has no finite poles, it is a polynomial; it is then constant because it has only a single zero (see Proposition 2.34 and Corollary 2.33). Then $\operatorname{div} r = 0$, and $T = \mathcal{O}$.

5. Let T be a fixed m-torsion point. We have to show that
$$\frac{g_{\alpha(T)} \circ \tau_{\alpha(S)}}{g_{\alpha(T)}} = \left(\frac{g_T \circ \tau_S}{g_T}\right)^{\deg \alpha}$$
for any m-torsion point S. The left hand side is a constant and hence can be composed with α without changing its value:
$$\begin{aligned}\frac{g_{\alpha(T)} \circ \tau_{\alpha(S)}}{g_{\alpha(T)}} &= \frac{g_{\alpha(T)} \circ \tau_{\alpha(S)} \circ \alpha}{g_{\alpha(T)} \circ \alpha} \\ &= \frac{g_{\alpha(T)} \circ \alpha \circ \tau_S}{g_{\alpha(T)} \circ \alpha}\end{aligned}$$
The right hand side is
$$\left(\frac{g_T \circ \tau_S}{g_T}\right)^{\deg \alpha} = \frac{g_T^{\deg \alpha} \circ \tau_S}{g_T^{\deg \alpha}}.$$
So we have to show that
$$\frac{g_{\alpha(T)} \circ \alpha}{g_T^{\deg \alpha}} \circ \tau_S = \frac{g_{\alpha(T)} \circ \alpha}{g_T^{\deg \alpha}} \quad \text{for any } m\text{-torsion point } S.$$
This is equivalent by Lemma 3.58 to
$$\frac{g_{\alpha(T)} \circ \alpha}{g_T^{\deg \alpha}} = r \circ [m]$$
for a suitable rational function r.
$$\begin{aligned}\operatorname{div}\left(\frac{g_{\alpha(T)} \circ \alpha}{g_T^{\deg \alpha}}\right) &= (\alpha^* \circ [m]^*)(\langle \alpha(T)\rangle - \langle \mathcal{O}\rangle) - \deg \alpha\, [m]^*(\langle T\rangle - \langle \mathcal{O}\rangle) \\ &= [m]^*\left(\alpha^*(\langle \alpha(T)\rangle - \langle \mathcal{O}\rangle) - \deg \alpha\, (\langle T\rangle - \langle \mathcal{O}\rangle)\right) \\ &\quad \text{since } \alpha \text{ and } [m] \text{ and thus } \alpha^* \text{ and } [m]^* \text{ commute} \\ &= [m]^*\left(\sum_{R \in \ker \alpha} e_\alpha\left(\langle T+R\rangle - \langle R\rangle\right) - \deg \alpha\left(\langle T\rangle - \langle \mathcal{O}\rangle\right)\right)\end{aligned}$$

It is now sufficient to verify that the divisor in large parentheses is principal. The degree of this divisor is obviously zero, and

$$\begin{aligned} \sum_{R \in \ker \alpha} e_\alpha (T + R - R) - \deg \alpha\, (T - \mathcal{O}) &= (e_\alpha |\ker \alpha| - \deg \alpha)\, T \\ &= \mathcal{O} \quad \text{by the definition of} \deg \alpha \end{aligned}$$

□

We can finally show that $e_m(S, T)$ is an m-th root of unity:

$$\begin{aligned} e_m(S,T)^m &= e_m(S, mT) \quad \text{by the bilinearity} \\ &= e_m(S, \mathcal{O}) \\ &= 1 \end{aligned}$$

3.8 HASSE'S THEOREM

We have now all ingredients for Hasse's famous theorem, which he proved in [Hasse, 1934] for odd characteristic:

Theorem 3.61 (Hasse) *Let $k = \mathbb{F}_q$ and $t = q + 1 - |E_k|$. Then the Frobenius endomorphism φ satisfies the following relations:*

1. $\varphi \circ \varphi - [t] \circ \varphi + [q] = [0]$
2. $|t| \leq 2\sqrt{q}$

Our proof is identical to that in [Charlap and Robbins, 1988], Section 12, pp. 69–72; we report it here for the sake of completeness. With the preparations of the previous sections it is mainly a matter of linear algebra. We exploit the fact that for a fixed integer m coprime to p, $E[m] \simeq \mathbb{Z}_m \times \mathbb{Z}_m$ is a free $\mathbb{Z}_m$-module of rank 2 (see Proposition 3.37).

Lemma 3.62 *Let $\{T_1, T_2\}$ be a basis for $E[m]$ as a $\mathbb{Z}_m$-module. Then the value $e_m(T_1, T_2)$ is a primitive m-th root of unity.*

Proof: Suppose that $e_m(T_1, T_2)^n = 1$. Then for $c_1, c_2 \in \mathbb{Z}_m$ we have:

$$\begin{aligned} e_m(nT_1, c_1T_1 + c_2T_2) &= e_m(T_1, c_1T_1 + c_2T_2)^n \text{ by Proposition 3.60, 1.} \\ &= e_m(T_1, T_1)^{nc_1}\, e_m(T_1, T_2)^{nc_2} \text{ by Proposition 3.60, 2.} \\ &= 1 \text{ by Proposition 3.60, 1.} \end{aligned}$$

Since $\{T_1, T_2\}$ is a basis, $c_1T_1 + c_2T_2$ varies over all of $E[m]$, and by the non-degeneracy property of e_m (Proposition 3.60, 4.), we conclude that $nT_1 = \mathcal{O}$, so $m|n$. □

The next theorem relates global information on an endomorphism with local information about the restriction of this endomorphism to $E[m]$. For its proof we use the Weil pairing.

Theorem 3.63 *Let α be a non-zero endomorphism. Then the restriction of α to $E[m]$, which we denote by α_m, is a linear endomorphism, and its determinant is* $\deg \alpha \pmod m$.

Proof: It is clear that $\alpha(E[m]) \subseteq E[m]$ and that α_m is linear. So α_m is a well defined endomorphism of $E[m]$ as a $\mathbb{Z}_m$-module. Choose a basis $\{T_1, T_2\}$ for $E[m]$. Then α_m can be represented by a matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

where $\alpha_m(T_j) = a_{1j}T_1 + a_{2j}T_2$. Then we compute

$$\begin{aligned} e_m(T_1, T_2)^{\deg \alpha} &= e_m(\alpha(T_1), \alpha(T_2)) \quad \text{by Proposition 3.60, 5.} \\ &= e_m(a_{11}T_1 + a_{21}T_2,\, a_{12}T_1 + a_{22}T_2) \\ &= e_m(T_1, T_1)^{a_{11}a_{12}} e_m(T_1, T_2)^{a_{11}a_{22}} e_m(T_2, T_1)^{a_{21}a_{12}} \\ &\quad\; e_m(T_2, T_2)^{a_{21}a_{22}} \\ &= e_m(T_1, T_2)^{a_{11}a_{22} - a_{21}a_{12}} \\ &= e_m(T_1, T_2)^{\det \alpha_m} \end{aligned}$$

As $e_m(T_1, T_2)$ is a primitive m-th root of unity by the lemma,

$$\deg \alpha = \det \alpha_m \pmod m.$$

□

Our next goal is to determine the degree of a linear combination of two endomorphisms.

Proposition 3.64 *Let α and β be non-zero endomorphisms, $c_1, c_2 \in \mathbb{Z}$. Then*

$$\deg([c_1] \circ \alpha + [c_2] \circ \beta) = c_1^2 \deg \alpha + c_2^2 \deg \beta + c_1 c_2 (\deg(\alpha + \beta) - \deg \alpha - \deg \beta)$$

Proof: Let m be an integer coprime to p and larger than both sides of the formula in the proposition. Restrict all endomorphisms to $E[m]$ and note that the structure of $\mathrm{End}(E)$ as $\mathbb{Z}$-module is compatible with that of $\mathrm{End}(E)|_{E[m]}$ as $\mathbb{Z}_m$-module via

$$([c] \circ \alpha)_m = c\alpha_m.$$

Then by Theorem 3.63,

$$\deg([c_1] \circ \alpha + [c_2] \circ \beta) = \det(c_1\alpha_m + c_2\beta_m) \pmod m.$$

Now it can be verified by a trivial computation on 2 × 2-matrices that

$$\begin{aligned} \det(c_1\alpha_m + c_2\beta_m) &= c_1^2 \det \alpha_m + c_2^2 \det \beta_m \\ &\quad + c_1c_2(\det(\alpha_m + \beta_m) - \det \alpha_m - \det \beta_m). \end{aligned}$$

Another application of Theorem 3.63 yields the desired result. □

Proposition 3.65 *If α is an endomorphism, then*

$$\beta := \alpha \circ \alpha - [1 + \deg \alpha - \deg([1] - \alpha)] \circ \alpha + [\deg \alpha] = [0]$$

Proof: We restrict as before all endomorphisms to $E[m]$ with m coprime to p. Then

$$\beta_m = \alpha_m^2 - (1 + \det \alpha_m - \det(\mathrm{id} - \alpha_m))\alpha_m + \det \alpha_m .$$

Another easy computation on 2×2-matrices yields

$$\mathrm{Tr}\, \alpha_m = 1 + \det \alpha_m - \det(\mathrm{id} - \alpha_m),$$

so that $\beta_m = 0$ by the Cayley–Hamilton Theorem. By varying m we find that $\beta(P) = \mathcal{O}$ for infinitely many torsion points P. So $\beta = [0]$ since otherwise $\beta = (\beta_1, \beta_2)$, where the β_i could have only a finite number of poles. □

Proof of Hasse's Theorem: Note that k is the fixed field of $x \mapsto x^q$, so

$$\begin{aligned} E_k &= \{(x,y) \in E : (x^q, y^q) = (x,y)\} \cup \{\mathcal{O}\} \\ &= \{P \in E : \varphi(P) = P\} \\ &= \ker([1] - \varphi). \end{aligned}$$

Since $[1] - \varphi$ is separable by Corollary 3.36,

$$\deg([1] - \varphi) = |\ker([1] - \varphi)| = |E_k|.$$

Finally $\deg \varphi = q$ by the example after Definition 3.20. The previous proposition, applied to $\alpha = \varphi$, yields the first part of the theorem.

For the second part we note that for any $c_1, c_2 \in \mathbb{Z}\backslash\{0\}$,

$$\begin{aligned} c_1^2 + c_2^2 q - c_1 c_2 t &= \deg([c_1] \circ [1] + [c_2] \circ (-\varphi)) \quad \text{by Proposition 3.64} \\ &> 0. \end{aligned}$$

Division by c_2^2 yields

$$\left(\frac{c_1}{c_2}\right)^2 - \frac{c_1}{c_2} t + q \geq 0$$

or

$$r^2 - rt + q \geq 0 \quad \text{for all } r \in \mathbb{Q}.$$

Then the inequality holds over $\mathbb{R}$, and the discriminant of the corresponding quadratic equation, $t^2 - 4q$, cannot be positive. Hence $t^2 \leq 4q$. □

3.9 WEIL'S THEOREM

In a special case the cardinality of an elliptic curve can be computed explicitly with little effort: If $L = \mathbb{F}_{q^m}$ is the extension of $k = \mathbb{F}_q$ of degree m and $|E_k|$ is known, there is a formula for $|E_L|$. In particular, if k is small, then $|E_k|$ can be determined by testing all possible X- and Y-coordinates if they result in a point on the curve. Such curves, like

$$Y^2 + XY = X^3 + X^2 + 1 \text{ over } \mathbb{F}_{2^m},$$

suggested in [Koblitz, 1991], p. 158, are actually used for cryptosystems.

Let $s = q+1-|E_k|$ and $t = q^m+1-|E_L|$. Then by Hasse's theorem we know that $\varphi_k : (x,y) \mapsto (x^q, y^q)$ is a zero in $\mathrm{End}(E)$ of the polynomial $T^2 - sT + q \in \mathbb{Z}[T]$. By the same argument t is an integer for which $\varphi_L : (x,y) \mapsto \left(x^{q^m}, y^{q^m}\right)$ is a zero of $T^2 - tT + q^m$. Moreover, φ_L satisfies no other equation $T^2 - t'T + q^m$ for $t' \neq t$. Otherwise, subtracting the equality $\varphi_L \circ \varphi_L + [q^m] = [t] \circ \varphi_L$ from $\varphi_L \circ \varphi_L + [q^m] = [t'] \circ \varphi_L$ yields $[0] = [t' - t] \circ \varphi$. Since φ is surjective by Proposition 3.3, it follows that $[t' - t] = [0]$ and $t' = t$. Exploiting the fact that $\varphi_L = \varphi_k^m$, we conclude that t is the unique integer for which φ_k is a zero of $T^{2m} - tT^m + q^m$.

To construct such a polynomial, we examine the complex zeros α and β of $T^2 - sT + q$. Since the discriminant $D = s^2 - 4q$ of this quadratic polynomial is not positive by Hasse's Theorem 3.61, α and β are complex conjugates. Moreover, they are integers in the imaginary quadratic number field $\mathbb{Q}(\sqrt{D})$, i.e. elements of $\mathbb{Z}[\vartheta]$ with $\vartheta = \sqrt{D}$ for $D \not\equiv 1 \pmod 4$ or $\vartheta = \frac{1+\sqrt{D}}{2}$ for $D \equiv 1 \pmod 4$. Consider the polynomial

$$f(T) = T^{2m} - (\alpha^m + \beta^m)T^m + q^m.$$

As α^m and β^m are complex conjugates, $\alpha^m + \beta^m$ is a real number. So $\alpha^m + \beta^m \in \mathbb{Z}[\vartheta] \cap \mathbb{R} = \mathbb{Z}$, and $f \in \mathbb{Z}[T]$. Furthermore, $f(\alpha) = -\alpha^m\beta^m + q^m = 0$ because $\alpha\beta = q$, and in the same way $f(\beta) = 0$. If $D < 0$, then α and β are distinct roots of f so that $T^2 - sT + q = (T - \alpha)(T - \beta)$ divides f in $\mathbb{Z}[T]$. Then φ_k must be a zero of f. If $D = 0$, we have to show that $\alpha = \beta$ is a double root of f, and the same argumentation holds. In this case

$$f'(T) = 2mT^{m-1}(T^m - \alpha^m)$$

has a zero in α, so f has at least a double zero in α.

We have thus proved the following theorem:

Theorem 3.66 (Weil) *Let E be defined over $\mathbb{F}_q$, $|E_{\mathbb{F}_q}| = q + 1 - s$ and m a positive integer. Decompose $T^2 - sT + q = (T - \alpha)(T - \beta)$ over the complex numbers. Then*

$$|E_{\mathbb{F}_{q^m}}| = q^m + 1 - (\alpha^m + \beta^m).$$

Example. $Y^2 + Y = X^3 + X + 1$ is defined over $\mathbb{F}_2$; replacing X and Y by the elements 0 and 1 of $\mathbb{F}_2$ we find that the equality never holds, so $E_{\mathbb{F}_2} = \{O\}$

and $s = 2$. We factor

$$T^2 - 2T + 2 = (T - (1+i))(T - (1-i)).$$

Hence

$$\begin{aligned} |E_{\mathbb{F}_{2^m}}| &= 2^m + 1 - ((1+i)^m + (1-i)^m) \\ &= 2^m + 1 - \begin{cases} 2 \cdot (-4)^{\frac{m}{4}} & \text{for } m \equiv 0 \pmod 4 \\ 0 & \text{for } m \equiv 2 \pmod 4 \\ 2 \cdot (-4)^{\frac{m-1}{4}} & \text{for } m \equiv 1 \pmod 4 \\ (-4)^{\frac{m+1}{4}} & \text{for } m \equiv 3 \pmod 4 \end{cases} \end{aligned}$$

For $m = 2$ we obtain $|E_{\mathbb{F}_4}| = 5$, a result which coincides with the example after Corollary 2.48.

3.10 TWISTED CURVES

If we know the cardinality of an elliptic curve E over $k = \mathbb{F}_q$, then we can easily define another curve E' whose cardinality is also known. Precisely, we construct E' such that there is a bijection between the X-coordinates of points on E_k and the elements of k which are *not* X-coordinates of points on E'_k. Then it turns out that $|E'_k| = q + 1 + t$ for $|E_k| = q + 1 - t$. In the first place we determine when a given $x \in k$ occurs as X-coordinate of a point on E_k, which amounts to solving the quadratic equation $E(x, Y)$ over k.

Proposition 3.67 *Let $Y^2 + aY + b \in k[Y]$ be a quadratic equation.*

1. *If $p \neq 2$, then the equation has its roots in k if and only if one of the following equivalent assertions holds:*
   - *$a^2 - 4b$ is zero or a quadratic residue in k.*
   - $\gcd(Y^2 + aY + b, Y^q - Y) \neq 1$
2. *If $p = 2$, denote by*

$$\mathrm{Tr} : k \to \mathbb{F}_2, \quad x \mapsto \sum_{i=0}^{m-1} x^{2^i}$$

the absolute trace function of k. Then the quadratic equation has its roots in k if and only if

$$a = 0 \quad or \quad \mathrm{Tr}(a^{-2}b) = 0.$$

Proof:

1. We recall that

$$\sigma : k^\times \to k^\times, \quad x \mapsto x^2$$

is a multiplicative group homomorphism with kernel $\{\pm 1\}$. So the image of σ is a subgroup of $k^\times$ of index 2. Its elements are called *quadratic residues*, while $k^\times \backslash \sigma(k^\times)$ consists of the *quadratic non-residues*. We introduce the quadratic character

$$\chi : k^\times \to \{\pm 1\}, \quad x \mapsto \begin{cases} 1 & \text{if } x \text{ is a quadratic residue} \\ -1 & \text{if } x \text{ is a quadratic non-residue} \end{cases}$$

Letting $\chi(0) = 0$ extends χ to a multiplicative function on k. Now completing the square

$$Y^2 + aY + b = \left(Y + \frac{a}{2}\right)^2 - \frac{a^2 - 4b}{4}$$

shows that the quadratic equation is solvable in k precisely if $\chi(a^2 - 4b) \neq -1$. More generally, the number of distinct solutions in k is $\chi(a^2 - 4b) + 1$. This proves the first condition.

In the most interesting case for implementational purposes, where p is a (large) prime, $\chi(x)$ is the Legendre symbol $\left(\frac{x}{p}\right)$ and can be evaluated efficiently by the Law of Quadratic Reciprocity, which was discovered by Gauß (see [Gauß, 1801], Article 131, or [Koblitz, 1994], Chapter II.2 for an elementary treatment). In the general case, however, we need a different characterisation. Taking into account that

$$Y^q - Y = \prod_{y \in k} (Y - y),$$

it is clear that

$$\gcd(Y^2 + aY + b, Y^q - Y) = \prod_{y \in k : y^2 + ay + b = 0} (Y - y),$$

which proves the second condition.

2. The assertion is trivial for $a = 0$ because $x \mapsto x^2$ is an automorphism of $k = \mathbb{F}_{2^m}$. Otherwise, we can apply the change of variables $Y \mapsto aY$ and divide the resulting equation by a^2 and hence assume without loss of generality that $a = 1$. Suppose that $y \in k$ is a root of $Y^2 + Y = b$. Repeated squaring of the equation yields

$$y^{2^i} + y^{2^{i-1}} = b^{2^{i-1}} \quad \text{for } 1 \leq i \leq m.$$

Summing up all these equations, we obtain

$$0 = y^{2^m} + y = \operatorname{Tr} b,$$

so the given condition is necessary.

Observe that if y is a solution, then $y + 1$ is the other one. Since the solutions come in pairs, the set of all equations $Y^2 + Y = b$ which have a root in k —

i.e. the set of $(Y+y)(Y+y+1)$ with $y \in k$ — forms one half of all these equations. On the other hand, the equations with $\operatorname{Tr} b = 0$, which comprise the equations solvable in k, are one half of all equations, too, so the two sets must coincide.

□

For the construction of E' we consider first the case $p \neq 2$ and suppose E in the normal form $Y^2 = s(X)$ with $s(X) = X^3 + a_2X^2 + a_4X + a_6$. Then it follows from the proof of the previous lemma that

$$\begin{aligned} |E_k| &= |E_k \backslash \{\mathcal{O}\}| + 1 \\ &= \sum_{x\in k} (\chi(4s(x)) + 1) + 1 \\ &= q + 1 + \chi(2)^2 \sum_{x \in k} \chi(s(x)) \\ &= q + 1 + t \end{aligned}$$

with $t = \sum_{x\in k} \chi(s(x))$.

Fixing a quadratic non-residue $\gamma \in k$, we define the elliptic curve $E' : Y^2 = s'(X)$ by $s'(X) = X^3 + \gamma a_2 X^2 + \gamma^2 a_4 X + \gamma^3 a_6$. Then $s'(\gamma x) = \gamma^3 s(x)$, and

$$\begin{aligned} |E'_k| &= q + 1 + \sum_{x\in k} \chi(s'(x)) \\ &= q + 1 + \sum_{x\in k} \chi(s'(\gamma x)) \\ &= q + 1 + \chi(\gamma)^3 \sum_{x\in k} \chi(s(x)) \\ &= q + 1 - t \end{aligned}$$

For $p = 2$ we consider the general elliptic curve $E : Y^2 + (a_1X + a_3)Y = s(X)$ with $s(X) = X^3 + a_2X^2 + a_4X + a_6$ and $a_1X + a_3 \neq 0$. The previous lemma implies that

$$\begin{aligned} |E_k| &= |E_k \backslash E[2]| + |E_k \cap E[2]| \\ &= 2\left|\left\{x \in k : a_1x + a_3 \neq 0,\ \operatorname{Tr}\left((a_1x + a_3)^{-2} s(x)\right) = 0\right\}\right| + |E_k \cap E[2]|. \end{aligned}$$

Recall from Section 2.5 that

$$E[2] \subseteq E_k$$

and

$$|E[2]| = \left\{ \begin{array}{ll} 1 & \text{for } a_1 = 0 \\ 2 & \text{for } a_1 \neq 0 \end{array} \right\} = |\{x \in k : a_1x + a_3 = 0\}| + 1.$$

Let $\gamma \in k$ be an element of trace 1, and let E' be the elliptic curve

$$Y^2 + (a_1X + a_3)Y = s'(X) \text{ with } s'(X) = s(X) + \gamma(a_1X + a_3)^2.$$

Then

$$\begin{aligned}\mathrm{Tr}\left((a_1x+a_3)^{-2}s'(x)\right) &= \mathrm{Tr}\left((a_1x+a_3)^{-2}s(x)+\gamma\right)\\ &= \mathrm{Tr}\left((a_1x+a_3)^{-2}s(x)\right)+1\end{aligned}$$

and

$$\begin{aligned}|E_k|+|E'_k| &= 2\left|\left\{x\in k: a_1x+a_3\neq 0,\ \mathrm{Tr}\left((a_1x+a_3)^{-2}s(x)\right)=0\right\}\right|\\ &\quad +2\left|\left\{x\in k: a_1x+a_3\neq 0,\ \mathrm{Tr}\left((a_1x+a_3)^{-2}s(x)\right)=1\right\}\right|\\ &\quad +2|E[2]|\\ &= 2|\{x\in k: a_1x+a_3\neq 0\}|+2\left(|\{x\in k: a_1x+a_3=0\}|+1\right)\\ &= 2(q+1)\end{aligned}$$

It follows that for $|E_k|=q+1+t$, we have $|E'_k|=q+1-t$.

We summarise our observations as follows:

Definition and Proposition 3.68 *Let E be an elliptic curve over k. Define another elliptic curve E' as follows:*

1. *If $p\neq 2$, E is of the form $Y^2=X^3+a_2X^2+a_4X+a_6$ and γ is a quadratic non-residue in k, then let*

$$E': Y^2=X^3+\gamma a_2X^2+\gamma^2a_4X+\gamma^3a_6.$$

2. *If $p=2$, E is of the form $Y^2+(a_1X+a_3)Y=X^3+a_2X^2+a_4X+a_6$ and γ is an element of k of trace 1, then let*

$$E': Y^2+(a_1X+a_3)Y=X^3+(a_2+\gamma a_1^2)X^2+a_4X+(a_6+\gamma a_3^2).$$

E' is called a twist *of E by γ. If $|E_k|=q+1+t$, then $|E'_k|=q+1-t$.*

The notion of a twist is symmetric in the sense that if E' is a twist of E (by γ), then E is a twist of E' (by γ^{-1} for $p\neq 2$ resp. γ for $p=2$). However, the twist of E is not unique: Since there are $\frac{q-1}{2}$ quadratic non-residues for $p\neq 2$ resp. $\frac{q}{2}$ elements of trace 1 for $p=2$, E has $\frac{q-1}{2}$ resp. $\frac{q}{2}$ different twists. But as the following proposition shows, this difference is not essential.

Proposition 3.69 *Let E be an elliptic curve over k. Then all the twists of E by quadratic non-residues resp. elements of trace 1 are isomorphic over k, thus twisting defines a bijection on the k-isomorphism classes of elliptic curves.*

Here "isomorphic over k" means that there is an admissible change of variables with coefficients in k which transforms one twist into another, see also Section 2.3.

Proof:

1. Let $p \neq 2$, $E : Y^2 = X^3 + a_2X^2 + a_4X + a_6$. Since two different twists of E by quadratic non-residues differ only by a twist by a quadratic residue, it is sufficient to show that if $\gamma = \delta^2$ is a quadratic residue in k, then $E'' : Y^2 = X^3 + \gamma a_2 X^2 + \gamma^2 a_4 X + \gamma^3 a_6$ is isomorphic to E. This is easy to verify because the admissible change of variables $(X, Y) \mapsto (\gamma^{-1}X, (\gamma\delta)^{-1}Y)$ transforms E to E''.

2. If $p = 2$ and $E : Y^2 + (a_1X + a_3)Y = X^3 + a_2X^2 + a_4X + a_6$, we have to show that for an element γ of trace zero, $E'' : Y^2 + (a_1X + a_3)Y = X^3 + (a_2 + \gamma a_1^2)X^2 + a_4X + (a_6 + \gamma a_3^2)$ is isomorphic to E.

   Suppose that s and t are solutions in k of the quadratic equations
   $$s^2 + a_1 s + \gamma a_1^2 = 0 \quad \text{and} \quad t^2 + a_3 t + \gamma a_3^2 = 0.$$
   Such solutions exist for $\operatorname{Tr}\gamma = 0$ by Proposition 3.67. The admissible change of variables $(X, Y) \mapsto (X, Y + sX + t)$ transforms E into $E'' + (a_3 s + a_1 t)X$, and we have to show that $a_3 s + a_1 t = 0$. From the defining equations of s and t it follows that $(a_3 s + a_1 t)^2 = a_1 a_3 (a_3 s + a_1 t)$. So if $a_3 s + a_1 t \neq 0$, then $a_3 s + a_1 t = a_1 a_3$. In this case replace s be the second root $s' = s + a_1$ of $X^2 + a_1 X + \gamma a_1^2$, so that $a_3 s' + a_1 t = (a_3 s + a_1 t) + a_1 a_3 = 0$.

□

3.11 SUPERSINGULAR CURVES

Supersingular curves are a special class of elliptic curves whose cardinalities and group structures are particularly easy to determine. In this section we collect interesting results on this special type of curves; the proofs of most of them have originally been obtained using the general theory of function fields and of abelian varieties and are beyond the scope of this presentation, so we refer the interested reader to the literature.

Definition 3.70 *An elliptic curve E is called* supersingular *if* $\operatorname{End}(E)$ *is non-commutative.*

Theorem 3.71 *In the case $p = 2$ or $p = 3$, E is supersingular if and only if $j(E) = 0$.*

Proof: See [Deuring, 1941], pp. 253 and 255. □

In his doctoral thesis, Waterhouse determined which numbers occur as cardinalities of elliptic curves over finite fields (see [Waterhouse, 1969], p. 536):

Theorem 3.72 (Waterhouse) *Let $k = \mathbb{F}_q = \mathbb{F}_{p^m}$ and t an integer such that $|t| \leq 2\sqrt{q}$. Then there is an elliptic curve E defined over k with $|E_k| = q + 1 - t$ if and only if one of the following conditions is satisfied:*

1. *p and t are coprime*
2. *m is even, and furthermore*
   - $t = \pm 2\sqrt{q}$
   - $t = \pm\sqrt{q}$ *and* $p \not\equiv 1 \pmod 3$, *or*
   - $t = 0$ *and* $p \not\equiv 1 \pmod 4$
3. *m is odd, and furthermore*
   - $t = 0$, *or*
   - $t = \pm\sqrt{pq}$ *and* $p = 2$ *or* 3

In the first case the curve is not supersingular, in the last two cases it is.

Corollary 3.73 *An elliptic curve E over a finite field $k = \mathbb{F}_q$ is supersingular if and only if $p \,|\, q + 1 - |E_k|$.*

The group structure of most of the supersingular curves can be derived directly from their cardinality, which was observed by Schoof (see [Schoof, 1987], p. 196):

Theorem 3.74 *Let E be a supersingular elliptic curve defined over the finite field $k = \mathbb{F}_q$, and $t = q+1-|E_k|$. Then the group structure of E_k is as follows:*

- *E_k is cyclic for $t^2 \in \{q, 2q, 3q\}$.*
- *$E_k \simeq \mathbb{Z}_{\sqrt{q}\pm 1} \times \mathbb{Z}_{\sqrt{q}\pm 1}$ for $t = \pm 2\sqrt{q}$*
- *E_k is cyclic for $t = 0$ and $q \not\equiv 1 \pmod 4$; it may be cyclic or of type $\mathbb{Z}_2 \times \mathbb{Z}_{\frac{q+1}{2}}$ for $t = 0$ and $q \equiv 1 \pmod 4$.*

In the same paper, Schoof gives the number $N(t)$ of non-isomorphic elliptic curves corresponding to the value t ([Schoof, 1987], pp. 194–195):

Theorem 3.75 (Schoof) *Let $k = \mathbb{F}_q = \mathbb{F}_{p^m}$ be a finite field, and denote by $N(t)$ the number of non-isomorphic elliptic curves E defined over k with $q + 1 - |E_k| = t$. Then for $|t| \leq 2\sqrt{q}$, $N(t)$ can be determined as follows:*

1. *$N(t) = H(t^2 - 4q)$ for t coprime to p*
2. *If m is even, then*

$$\begin{aligned} N(\pm 2\sqrt{q}) &= \frac{1}{12}\left(p + 6 - 4\left(\frac{-3}{p}\right) - 3\left(\frac{-4}{p}\right)\right) \\ N(\pm\sqrt{q}) &= 1 - \left(\frac{-3}{p}\right) \\ N(0) &= 1 - \left(\frac{-4}{p}\right) \end{aligned}$$

3. If m is odd, then

$$\begin{aligned} N(0) &= H(-4p) \\ N(\pm\sqrt{pq}) &= 1 \textit{ for } p = 2 \textit{ or } 3 \end{aligned}$$

Here $H(\Delta)$ denotes the Kronecker class number symbol for negative Δ (for a definition, see [Schoof, 1987], Section 2) and $\left(\frac{a}{p}\right)$ the Legendre symbol. It should be noted that the Kronecker symbol can only be evaluated for small values of $|\Delta|$; so the theorem is not suited for determining the number of isomorphism classes of non-supersingular curves over large fields. For supersingular curves in characteristic 2 or 3, the interesting values of $H(-4p)$ are $H(-8) = 1$ and $H(-12) = 2$.

For even characteristic, Menezes and Vanstone gave an elementary proof of the above theorem and moreover determined representatives for each isomorphism class (see [Menezes and Vanstone, 1990]). Table 3.1 summarises the results for supersingular curves and m even; γ is a non-cube, ω, α, β and δ are elements of $\mathbb{F}_{2^m}$ such that $\operatorname{Tr}\omega = \operatorname{Tr}(\gamma^{-2}\alpha) = Tr(\gamma^{-4}\beta) = 1$ and $\operatorname{Tr}_{\mathbb{F}_4}\delta \neq 0$, where $\operatorname{Tr}\kappa = \sum_{i=0}^{m-1}\kappa^{2^i}$ is the absolute trace of κ and $\operatorname{Tr}_{\mathbb{F}_4}\kappa = \sum_{i=0}^{m/2-1}\kappa^{4^i}$ the trace over the subfield $\mathbb{F}_4$. In the columns for t the upper sign is valid for $\frac{m}{2}$ even, the lower one for $\frac{m}{2}$ odd. Table 3.2 gives the representatives for m odd; the upper sign is valid for $m \equiv \pm 1 \pmod 8$, the lower one for $m \equiv \pm 3$ (mod 8).

In a similar approach, Morain classified the supersingular curves in characteristic 3. Tables 3.3 and 3.4 summarise the results from [Morain, 1997] for m even and odd. γ is a quadratic non-residue, δ an element whose absolute trace is 1. The upper sign stands for the case $\frac{m}{2}$ even (resp. $\frac{m-1}{2}$ even), the lower sign for $\frac{m}{2}$ odd (resp. $\frac{m-1}{2}$ odd).

representative	t
$Y^2 + \gamma Y = X^3$	$\pm\sqrt{q}$
$Y^2 + \gamma Y = X^3 + \alpha$	$\mp\sqrt{q}$
$Y^2 + \gamma^2 Y = X^3$	$\pm\sqrt{q}$
$Y^2 + \gamma^2 Y = X^3 + \beta$	$\mp\sqrt{q}$
$Y^2 + Y = X^3 + \delta X$	0
$Y^2 + Y = X^3$	$\mp 2\sqrt{q}$
$Y^2 + Y = X^3 + \omega$	$\pm 2\sqrt{q}$

Table 3.1. Isomorphism classes of supersingular curves over $\mathbb{F}_{2^m}$, m even

representative	t
$Y^2 + Y = X^3$	0
$Y^2 + Y = X^3 + X$	$\pm\sqrt{2q}$
$Y^2 + Y = X^3 + X + 1$	$\mp\sqrt{2q}$

Table 3.2. Isomorphism classes of supersingular curves over $\mathbb{F}_{2^m}$, m odd

representative	t
$Y^2 = X^3 - X$	$\pm 2\sqrt{q}$
$Y^2 = X^3 - \gamma^2 X$	$\mp 2\sqrt{q}$
$Y^2 = X^3 - \gamma X$	0
$Y^2 = X^3 - \gamma^3 X$	0
$Y^2 = X^3 - X + \delta$	$\mp\sqrt{q}$
$Y^2 = X^3 - \gamma^2 X + \gamma^3 \delta$	$\pm\sqrt{q}$

Table 3.3. Isomorphism classes of supersingular curves over $\mathbb{F}_{3^m}$, m even

representative	t
$Y^2 = X^3 + X$	0
$Y^2 = X^3 - X$	0
$Y^2 = X^3 - X + \delta$	$\mp\sqrt{3q}$
$Y^2 = X^3 - X - \delta$	$\pm\sqrt{3q}$

Table 3.4. Isomorphism classes of supersingular curves over $\mathbb{F}_{3^m}$, m odd

3.12 GROUP STRUCTURE

Once the cardinality of an elliptic curve E over $k = \mathbb{F}_q$ is determined, its group structure is of interest. By the Fundamental Theorem on Abelian Groups 3.47, we know that E_k is isomorphic to a unique direct product of cyclic groups

$$\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_r}$$

with $n_1 > 1$ and $n_i | n_{i+1}$ for $i = 1, \ldots r - 1$. Since E_k is finite and thus a torsion group, there is an integer m such that $E_k \subsetneq E[m]$. By Theorem 3.39, $E[m]$ is isomorphic to $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2}$ for suitable integers m_1 and m_2, whence $r \leq 2$ (see also Proposition 4.2).

This simple observation can be sharpened to the following theorem first proved in [Rück, 1987]:

Theorem 3.76 (Rück) *Let E be an elliptic curve defined over $\mathbb{F}_q$. Then*

$$E_{\mathbb{F}_q} \simeq \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2}$$

with $n_1 | n_2$ and $n_1 | q - 1$.

Note that this formulation covers the case of a cyclic group as well by letting $n_1 = 1$.

Proof: It remains to show that $n_1 | q - 1$. The following elementary proof was pointed out to me by Berit Skjernaa and Scott Vanstone. From $n_1 | n_2$ we conclude that E_k contains the subgroup $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_1}$ and thus n_1^2 points of n_1-torsion. By Theorem 3.39, these are all the n_1-torsion points.

We now use the Weil pairing to show that k contains a primitive n_1-th root of unity. By Lemma 3.62 and the above observation, there are points P, $Q \in E[n_1] \subsetneq E_k$ such that $e_{n_1}(P, Q)$ is a primitive n_1-th root of unity. Let φ be the Frobenius endomorphism of E_k. Then $\varphi(P) = P$ and $\varphi(Q) = Q$ imply

$$\begin{aligned} e_{n_1}(P, Q) &= e_{n_1}(\varphi(P), \varphi(Q)) \\ &= e_{n_1}(P, Q)^{\deg \varphi} \text{ by Proposition 3.60, 5.} \\ &= e_{n_1}(P, Q)^q \text{ by the example after Definition 3.20.} \end{aligned}$$

Hence, $e_{n_1}(P, Q) \in k$, and n_1, the multiplicative order of $e_{n_1}(P, Q)$, must divide $q - 1$, the order of $k^\times$. □

4 THE DISCRETE LOGARITHM PROBLEM

Откройте мне только вашу тайну.

—Пушкин

The public key cryptosystems presented in Chapter 1 rely on the difficulty of solving the discrete logarithm problem in certain groups: An adversary who could efficiently compute discrete logarithms in the group underlying such a cryptosystem would be able to break the system. So to judge the security of the proposed cryptosystems we must have a closer look at algorithms for solving discrete logarithm problems.

To provide a common framework for the following sections, we reformulate the problem and fix some notations: Let $G = \langle \alpha \rangle$ be a finite, multiplicatively written cyclic group with generator α and known cardinality n and let β be an element of G. The discrete logarithm problem is to compute an integer l (which is denoted by $\log_\alpha \beta$) such that $\beta = \alpha^l$. The integer l is determined uniquely modulo n. The problem can on one hand be solved by a *generic* or *black box algorithm*, which does not take into account the representation of group elements. We only require that it be possible to efficiently multiply and invert group elements and to test them for equality. We then solve the problem by these elementary operations, starting with the given elements α and β. The third requirement may seem surprising since in most groups it is easy to test whether two elements are equal; but it can be an issue in factor groups, which are given modulo an equivalence relation, so that the same element may have different representatives. An example is provided by the divisor class group of an elliptic curve, where the problem is solved by working with the unique representatives given by single points. Other examples are class groups of number fields or divisor class groups of more general curves, in which cases this issue is more serious.

It turns out, however, that the difficulty of the discrete logarithm problem depends heavily on the representation of the group. For instance, it is trivial for $G = \mathbb{Z}_n$ and $\alpha = 1$. More generally it is easy to solve for $G = \mathbb{Z}_n$ and any generator α of $\mathbb{Z}_n$ by the Euclidian algorithm. (Indeed, the discrete logarithm problem in G can be reformulated to the task of computing an explicit isomorphism of G with $\mathbb{Z}_n$.) Hence it is worthwhile to take the concrete representation of the group into account when looking for an efficient solution to the discrete logarithm problem. We will see in Section 4.4, for instance, that there are especially good algorithms for the multiplicative groups of finite fields. Some elliptic curves are also cryptographically insecure, which we will show in Section 4.5, using the preparations made in the previous chapter.

4.1 SHANKS'S BABY-STEP GIANT-STEP ALGORITHM

Shanks's baby-step giant-step technique is an (almost) generic algorithm; it basically amounts to testing all powers of α if they equal β. Using storage space and a suitable data structure, however, it is more efficient than this naïve approach. The algorithm has been introduced in [Shanks, 1971] to compute the class number of imaginary quadratic number fields. Variants of it can be used to determine the structure of different abelian groups. For instance, we will study an algorithm for computing the cardinality of an elliptic curve in Section 5.1.

The basic idea is as follows: Instead of computing α, α^2, ... until β is found, a smaller list of powers of α is precomputed ("baby-steps"). Then the powers of α are scanned in larger steps by computing α^b, α^{2b}, ... for a suitable positive integer b until an α^{ib} equals an element in the precomputed list ("giant-steps"). If α^{ib} is looked up in the baby-step list by simply comparing with each of the list entries, this algorithm has no advantage over the naïve one. However, if each group element has a unique representative and these representatives can be ordered, then it is possible to maintain the baby-steps in an ordered list, and the look-up can be performed efficiently by a binary search. This extra requirement theoretically makes the algorithm non-generic, but in practice it covers most of the interesting groups.

In the case of a finite prime field $\mathbb{F}_p^\times$ any element has a unique representative in the set $\{1, \dots, p-1\}$ and the natural order can be used. In $\mathbb{F}_{p^m}^\times$ any element can be represented by a unique element of $\mathbb{F}_p[X]$ of degree less than m, and the usual order by the degree in the first and the list of coefficients in the second place can be chosen. Finally in the case of an elliptic curve the points are elements of $\mathbb{F}_q \times \mathbb{F}_q$ and can be ordered lexicographically by the two components.

In detail the algorithm is as follows:

1. Fix a number of baby-steps b.
2. Compute the pairs (α^i, i) for $0 \leq i < b$ and store them in a list which is sorted by the first component.
3. The number of giant steps is $g = \lceil \frac{n}{b} \rceil$.
4. Compute $\beta\alpha^{-kb}$ for $0 \leq k < g$ and by a binary search look up $\beta\alpha^{-kb}$ in the table of precomputed α^i. If it is found, then $\log_\alpha \beta = kb + i$.

Note that the algorithm terminates because $\log_\alpha \beta$ is an element of the set $\{0, \ldots, n-1\}$ and hence has a unique representation

$$\log_\alpha \beta = kb + i \quad \text{with } 0 \leq i < b,\ 0 \leq k < \left\lceil \frac{n}{b} \right\rceil$$

Computing the baby steps and sorting the list takes $O(b \log b)$ operations, where an operation is a multiplication or a comparison of two group elements. The giant steps require $O(g \log b)$ operations, and the exact coefficients of both O's are of roughly equal size. Hence to determine a nearly optimal number of baby-steps we neglect the term $\log b$ and minimise $b + g$ subject to $bg = n$. The solution is $b = g = \sqrt{n}$, and by the necessary rounding we obtain $b = g = \lceil \sqrt{n} \rceil$. So the total time complexity is $O(\sqrt{n} \log n)$, and the space complexity is $O(\sqrt{n} \log n)$ to store $\lceil \sqrt{n} \rceil$ entries of length $O(\log n)$.

It should be noted that Shanks's algorithm is applicable even if the group order n is not known. First, it is sufficient to work with an upper bound for n. If not even an upper bound is achievable, then one can simply choose any n and, if the discrete logarithm is not found, repeat the algorithm with increasing values of n.

4.2 POLLARD'S ρ-METHOD

A drawback of Shanks's algorithm is the need for much memory when the group becomes large. Pollard suggested a probabilistic algorithm with approximately the same running time, but practically no memory requirements. The algorithm in [Pollard, 1978] is described for the multiplicative group of a finite prime field, but generalises immediately to arbitrary groups as follows.

Let $G = T_1 \dot{\cup} T_2 \dot{\cup} T_3$ be a random partition of G into three sets of roughly equal size. Select a group element $x_0 = \alpha^{a_0}\beta^{b_0}$ with random numbers a_0 and b_0, and construct recursively sequences $(x_i)_{i \geq 0}$, $(a_i)_{i \geq 0}$ and $(b_i)_{i \geq 0}$ by

$$x_{i+1} = \begin{cases} \beta x_i & \text{for } x_i \in T_1 \\ x_i^2 & \text{for } x_i \in T_2 \\ \alpha x_i & \text{for } x_i \in T_3 \end{cases}$$

$$a_{i+1} = \begin{cases} a_i & \text{for } x_i \in T_1 \\ 2a_i & \text{for } x_i \in T_2 \\ a_i + 1 & \text{for } x_i \in T_3 \end{cases} \qquad b_{i+1} = \begin{cases} b_i + 1 & \text{for } x_i \in T_1 \\ 2b_i & \text{for } x_i \in T_2 \\ b_i & \text{for } x_i \in T_3 \end{cases}$$

Then $x_i = \alpha^{a_i}\beta^{b_i}$ holds for all $i \geq 0$.

Since G is finite, the sequence (x_i) ultimately becomes periodic, i.e. there are unique smallest integers $\mu \geq 0$ (called the *preperiod*) and $\lambda \geq 1$ (called the *period*) such that $x_1, \ldots, x_{\mu+\lambda-1}$ are distinct and $x_{i+\lambda} = x_i$ for $i \geq \mu$. Drawing the elements of the sequence as points in the plane and joining x_i and x_{i+1} by a line results in a figure which reminds of the Greek letter ρ (see Figure 4.1), hence the naming.

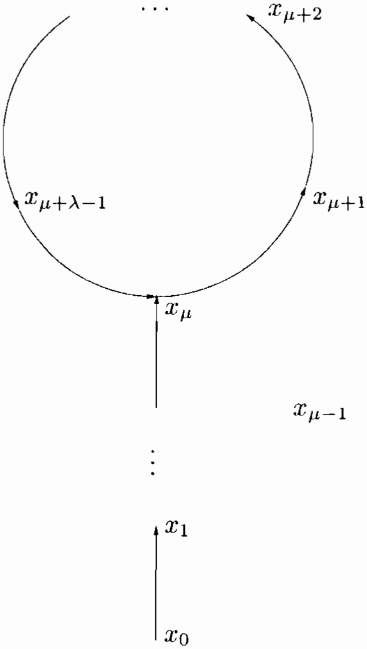

Figure 4.1. Pollard's ρ

The essence of the ρ-method is to find a match $x_i = x_j$ for $i \neq j$; then

$$\alpha^{a_i+lb_i} = \alpha^{a_i}\beta^{b_i} = x_i = x_j = \alpha^{a_j}\beta^{b_j} = \alpha^{a_j+lb_j}$$

implies

$$l(b_j - b_i) \equiv a_i - a_j \pmod{n}. \tag{4.1}$$

If $d = \gcd(n, b_j - b_i) = 1$, then this equation can be solved for l; otherwise, there are d possible values for l, which can be tested for $\alpha^l = \beta$ at least if d is small. Precisely, compute

$$d = un + v(b_j - b_i)$$

with integers u and v by the extended Euclidian algorithm. Multiplying (4.1) by v results in

$$ld \equiv v(a_i - a_j) \pmod{n}.$$

Equation (4.1) implies that $d|a_i - a_j$, *thus* l is one of the d distinct values

$$\frac{v(a_i - a_j)}{d} + k\frac{n}{d} \text{ for } 0 \leq k < d$$

modulo n. (In practice, the Pohlig–Hellman algorithm described in Section 4.3 allows to reduce the discrete logarithm problem to the case where n is prime. It is then extremely unlikely that d becomes big.)

The first match which can possibly be detected is $x_\mu = x_{\mu+\lambda}$, so that $\mu + \lambda$ sequence entries have to be computed. Assuming that the map

$$F : G \to G, \quad x \mapsto \begin{cases} \beta x_i & \text{for } x_i \in T_1 \\ x_i^2 & \text{for } x_i \in T_2 \\ \alpha x_i & \text{for } x_i \in T_3 \end{cases}$$

behaves like a random mapping, i.e. one which is chosen according to a uniform distribution over all maps $G \to G$, the expected value for $\mu + \lambda$ is close to

$$\sqrt{\frac{\pi}{2}n} \approx 1.25\sqrt{n} \in O(\sqrt{n}),$$

see [Teske, 1998].

It turned out in extensive experiments that Pollard's iterating function behaved worse than a true random mapping by a constant factor of about 1.2; however, different functions, which could be proved to asymptotically behave like a random walk, yielded the predicted complexity ([Teske, 1998]).

The space complexity of Pollard's algorithm depends on the implementation of the match finding procedure. The naïve approach of storing the sequence entries (x_i, a_i, b_i) in a structure sorted by the first component (for instance in a balanced tree) and looking up later entries needs to store $O(\sqrt{n})$ elements as Shanks's method. Pollard suggested a cycle finding technique due to Floyd (see [Knuth, 1981], Exercise 3.1.6), which consists of computing $(x_i, a_i, b_i, x_{2i}, a_{2i}, b_{2i})$ until $x_i = x_{2i}$. This happens as soon as i is a multiple of λ and not less than μ. Practically no storage space is required by this approach, since x_{i+1} and $x_{2(i+1)}$ can be computed from the previous values x_i and x_{2i} by three applications of the iterating function via $x_{i+1} = F(x_i)$ and $x_{2(i+1)} = F(F(x_{2i}))$. Since Floyd's algorithm usually does not detect the first possible match, the expected value of i for which a match is detected is higher than the expected value of λ; for a true random mapping it is

$$\frac{\pi^2}{12}\sqrt{\frac{\pi}{2}n} \approx 1.03\sqrt{n}.$$

Moreover, the values $x_1, \ldots, x_i$ are computed twice to obtain the sequences (x_i) and (x_{2i}), so that the total number of function evaluations is on average

$$3\frac{\pi^2}{12}\sqrt{\frac{\pi}{2}n} \approx 3.09\sqrt{n}.$$

More efficient algorithms for cycle detection are described in [Brent, 1980] and [Schnorr and Lenstra Jr., 1984], Section 3.

The ρ-method can be parallelised by starting the iterations on different machines and reporting the values obtained to a central server, which stores them and searches them for collisions. However, if all computed group elements were stored, then the time saved by the parallelisation would be more than compensated by the extra time needed on the central machine for the look-up of matches. It was suggested in [Oorschot and Wiener, 1999] to choose an easily recognisable set of *distinguished group elements*, for instance all elements whose binary representations start with some consecutive zeros. The iterations are performed until a distinguished point is reached, which is sent to the central server, then the process is repeated with a new starting point. If this algorithm runs on M machines in parallel, its running time is reduced by almost a factor of M compared to the serial ρ-method.

4.3 POHLIG–HELLMAN METHOD

If the group order n is not prime, then it is possible to take advantage of the factorisation of n by breaking the discrete logarithm problem in G into a number of discrete logarithm problems in the Sylow groups of G, which are the maximal subgroups of prime power order. Moreover, in these groups the problem can be solved by iteratively taking logarithms in the subgroup of the corresponding prime order. Hence the following algorithm, which has been described in [Pohlig and Hellman, 1978] and has independently been discovered by Roland Silver, and Richard Schroeppel and H. Block, runs in about $O(\sqrt{p}\log p)$ if p is the largest prime factor of n.

Write the decomposition of n into prime powers

$$n = \prod_{i=1}^{k} p_i^{\nu_i}.$$

If all of the p_i are small or only one of them is large, then the factorisation of n is easy to compute; if n has several large prime factors, however, it may be computationally infeasible to factor it. But in this case the discrete logarithm problems in the subgroups of (large) prime order are not solvable, anyway, so in practice this is no restriction on the applicability of the algorithm.

The discrete logarithm $l = \log_\alpha \beta$ is determined separately modulo all $p_i^{\nu_i}$, and the results are combined via the Chinese Remainder Theorem.

To compute $l \bmod p^\nu$ suppose that it is written to the base p as

$$l \equiv \sum_{i=0}^{\nu-1} b_i \, p^i \mod p^\nu.$$

We determine the b_i inductively. Let $\gamma = \alpha^{\frac{n}{p}}$ be a generator of the subgroup G_p of order p in G. Then

$$\beta^{\frac{n}{p}} = \alpha^{x\frac{n}{p}} = \gamma^x = \gamma^{b_0}.$$

Now b_0 is the solution of the discrete logarithm problem

$$b_0 = \log_\gamma \left(\beta^{\frac{n}{p}}\right)$$

in G_p, which can be solved by one of the methods of Sections 4.1 or 4.2. For the further coefficients assume that $b_0, \dots, b_{j-1}$ are already computed, and define l_j by

$$l_j = \sum_{i=0}^{j-1} b_i \, p^i.$$

Then

$$\left(\beta\alpha^{-l_j}\right)^{\frac{n}{p^{j+1}}} = \alpha^{\left(\sum_{i=j}^{\nu-1} b_i p^{i-j}\right)\frac{n}{p}} = \gamma^{b_j},$$

and

$$b_j = \log_\gamma \left(\left(\beta\alpha^{-x_j}\right)^{\frac{n}{p^{j+1}}}\right)$$

is again the solution of a discrete logarithm problem in G_p.

The computation of $(\beta\alpha^{-x_j})^{\frac{n}{p^{j+1}}}$ by the Square and Multiply Algorithm 1.1 requires $O(\log n)$ multiplications and the discrete logarithm in G_p is computed with $O(\sqrt{p}\log p)$ operations. So if p is the largest prime factor and $r = \sum_{i=1}^{k} \nu_i$ the total number of primes in the factorisation of n, the effort for determining x modulo all the prime powers is in

$$O\left(\sum_{i=1}^{k} \nu_i(\log n + \sqrt{p_i}\log p_i)\right) \subseteq O\left(r(\log n + \sqrt{p}\log p)\right).$$

The time needed for the combination of the results by the Chinese Remainder Theorem is in $O(r\log^2 n)$ *bit* operations (see [Cohen, 1993], pp. 12–20); since a group element is represented by at least $\log_2 n$ bits, any group operation takes at least $\log_2 n$ bit operations, and the overall complexity of the algorithm remains unchanged by this last step.

4.4 INDEX CALCULUS METHODS

By choosing n a large prime and noting that the input size of the discrete logarithm problem is in $\Omega(\log n)$, it is easy to see that the algorithms presented so far are fully exponential. This is not incidental: Shoup showed in [Shoup, 1997] that any algorithm for computing discrete logarithms in an arbitrary group requires $\Omega(\sqrt{p}\log p)$ steps, where p is the largest prime factor of n. So in the generic setting one cannot hope to improve substantially on the Pohlig–Hellman method presented in the previous section. More efficient algorithms can only be developed for specific groups, exploiting additional structure. In this section we treat a probabilistic subexponential algorithm which is adapted to the multiplicative group of a finite field.

Definition 4.1 *A (non-deterministic) algorithm with input size* $\log n$ *is* subexponential *if there are constants* $c > 0$ *and* $\alpha \in [0, 1)$ *so that the (expected) running time of the algorithm is in*

$$L[\alpha, c] = O\left(e^{(c+o(1))(\log n)^{\alpha}(\log\log n)^{1-\alpha}}\right).$$

Note that for $\alpha = 0$ a subexponential algorithm is polynomial, for $\alpha = 1$ it is fully exponential.

The algorithms presented in this section consist of two phases:

Phase 1: Collecting linear equations

Fix a *factor base* $\Gamma = \{\gamma_1, \ldots, \gamma_t\}$ of elements of G. In this first phase we try to determine the discrete logarithms of the γ_i. We repeatedly choose random integers $s \in \{0, \ldots, n-1\}$ and compute α^s, which we attempt to factor in Γ. If we are successful, the equation

$$\alpha^s = \prod_{i=1}^{t} \gamma_i^{\nu_i}$$

yields the linear equation in $\mathbb{Z}_n$

$$s = \sum_{i=1}^{t} \nu_i \log_\alpha \gamma_i.$$

When sufficiently many of these equations are collected they can be solved for the $\log_\alpha \gamma_i$'s.

Phase 2: Computing individual logarithms

Again we select random integers s and this time try to factor $\beta\alpha^{-s}$ in Γ. If we are successful and

$$\beta\alpha^{-s} = \prod_{i=1}^{t} \gamma_i^{\nu_i},$$

then

$$\log_\alpha \beta = s + \sum_{i=1}^{t} \nu_i \log_\alpha \gamma_i,$$

and all of the numbers on the right hand side are known.

Note that the first phase is required only once to compute arbitrarily many logarithms in the second phase. So there is a trade-off between the time expended for the two phases. Choosing a larger factor base slows down the first phase, but speeds up the second phase. Hence the more discrete logarithms have to be computed in the same field, the larger the factor base should be.

The applicability of the algorithm depends on whether or not a factor base can be chosen which allows to efficiently construct the relations above. To date suitable factor bases are known for finite fields and class groups of imaginary quadratic number fields ([McCurley, 1989]) and for divisor class groups of hyperelliptic curves of high genus ([Adleman et al., 1994], [Müller et al., 1997] and [Enge, 1998]).

In the case of a finite prime field $\mathbb{F}_p$ it is natural to construct the factor base from the smallest prime numbers and to represent the group elements by integers in the range $\{0, \dots, p-1\}$. A group element γ is factored within the factor base by *trial division*, i.e. factors γ_1 are split from γ until the resulting number is no more divisible by γ_1, then the procedure is continued with γ_2 and so on. If $r \leq \log_2 p$ is the total number of (not necessarily distinct) primes in the factorisation of γ, then at most $t + r$ trial divisions by elements of Γ have been sufficient to factor γ within Γ or to prove that such a factorisation does not exist.

The elements of $\mathbb{F}_{p^m}$ can be represented by polynomials in $\mathbb{F}_p[X]$ of degree less than m, and an obvious choice for Γ are all irreducible polynomials of small degree.

The basic concept has known a great variety of improvements and has led to different implementations. Blake, Fuji-Hara, Mullin and Vanstone introduced the technique of "systematic equations", finding that a large portion of the equations needed in the first phase can be obtained almost for free ([Blake et al., 1984]). Using an algorithm due to Coppersmith in connection with a polynomial sieve, Gordon and McCurley were able to compute logarithms in $\mathbb{F}_{2^{401}}$. They claim that "computing discrete logarithms in GF (2^{503}) might be possible within the next 5–10 years" ([Gordon and McCurley, 1993], p. 322). Prime fields are apparently harder to attack; the current record is for a prime of 65 digits ([Weber, 1996]).

The algorithms with the best rigorously proven running time for $\mathbb{F}_p$ and $\mathbb{F}_{2^m}$ are due to Pomerance; they run in $L[1/2, \sqrt{2}]$ ([Pomerance, 1987]). It is unknown whether a subexponential algorithm exists for $\mathbb{F}_{p^m}$ where both p and m vary. Lovorn Bender was able to prove the subexponentiality for $\mathbb{F}_{p^2}$, see [Lovorn Bender, 1999]. There are various algorithms with a conjectured, but unproven better running time. An index calculus method for $\mathbb{F}_p$, which is based on the *number field sieve*, has a running time of $L[1/3, 4/\sqrt[3]{9}]$, see [Gordon, 1993] and [Schirokauer, 1993]. Coppersmith presented an algorithm for $\mathbb{F}_{2^m}$, which can be interpreted as a special case of the number field sieve, with a running time of $L[1/3, c]$ where c is about 1.4 ([Coppersmith, 1984]). These two algorithms were used to obtain the records mentioned above. An analogue to the number field sieve over function fields, the *function field sieve*, was introduced in [Adleman, 1994]; its running time for $\mathbb{F}_{2^m}$ is $L[1/3, \sqrt[3]{9}]$.

In view of these results cryptosystems based on the discrete logarithm problem in a finite field are generally considered secure if the field size has about 1000 bits.

It should be noted that collecting linear equations in the first phase as well as the complete second phase are easily parallelisable. The bottleneck in all implementations is the solution of the huge, but sparse system of linear equations over $\mathbb{Z}_n$ obtained in the first phase. An efficient parallel implementation would push the applicability of the index calculus method much further.

4.5 ELLIPTIC CURVE LOGARITHMS

The big advantage of elliptic curve cryptosystems over those based on the discrete logarithm problem in finite fields is that for the former no subexponential algorithms are known, with the exception of some rare classes of curves.

Index calculus

Miller argued that an analogue of the index calculus method is improbable to exist for elliptic curves, and his arguments are confirmed by further theoretical considerations and computational experiments by Silverman and Suzuki. We give a rough outline of the arguments and refer the reader interested in a more rigorous reasoning to [Miller, 1986] and [Silverman and Suzuki, 1998]. If the elliptic curve E is defined over the finite prime field $\mathbb{F}_p$, then its coefficients can be lifted to integers, and we can consider one of the possible lifts $E_{\mathbb{Q}}$. (For the second important class of elliptic curves, those defined over $\mathbb{F}_{2^m}$, the coefficients must be lifted to integers in a number field of degree m in which 2 is not inert, and a similar argumentation applies.) Restricting to $E_{\mathbb{Z}_{(p)}}$, where

$$\mathbb{Z}_{(p)} = \left\{ \frac{a}{b} \in \mathbb{Q} : p \nmid b \right\}$$

is the localised ring of $\mathbb{Z}$ at p, the coordinates of a point on $E_{\mathbb{Z}_{(p)}}$ can be reduced modulo p to obtain a point on $E_{\mathbb{F}_p}$. The index calculus method would then work as follows. Fix a factor base containing some points on $E_{\mathbb{F}_p}$, which are lifted to points on $E_{\mathbb{Z}_{(p)}}$. Choose random elements on $E_{\mathbb{F}_p}$ as described in the previous section, and "factor" them by lifting them to points on $E_{\mathbb{Z}_{(p)}}$ and expressing them as a linear combination of the lifted factor base. Then the reduced points satisfy the same linear relation on $E_{\mathbb{F}_p}$.

However, no simple procedure is known for lifting points on $E_{\mathbb{F}_p}$ to $E_{\mathbb{Z}_{(p)}}$. Furthermore, the whole algorithm can only be efficient if the factor base consists of points of small height, where the *logarithmic height* of a point $\left(\frac{x}{d}, \frac{y}{d}\right)$ with x, y, $d \in \mathbb{Z}$ and $\gcd(x, y, d) = 1$ is defined by $\log \max\{|x|, |y|, |d|\}$, and hence is up to a constant factor the number of digits needed to represent the point. But the logarithmic heights of the multiples nP of a point P grow very fast (namely quadratically in n), as is illustrated by Figure 4.2, which records the X-coordinates of successive multiples of $P = (1,1)$ on the rational curve $Y^2 = X^3 + X - 1$, starting with $4P$. The figure is inspired by a similar one on page 143 of [Koblitz, 1998].

On the other hand, the maximal number of linearly independent points on a curve $E_{\mathbb{Q}}$, called its *rank*, is finite by the Mordell–Weil Theorem (see [Husemöller, 1987], Chapter 6, [Lang, 1978], Chapter IV.2 or [Silverman, 1986],

$$\frac{25}{36}$$

$$\frac{685}{121}$$

$$\frac{7082}{2209}$$

$$\frac{154513}{196249}$$

$$\frac{9781441}{197136}$$

$$\frac{645430801}{468073225}$$

$$\frac{54088691834}{39890874529}$$

$$\frac{23545957758733}{430654875049}$$

$$\frac{2660536742331673}{3348618159624516}$$

$$\frac{3438505996705270765}{1099386223759472401}$$

$$\frac{2389279734043328028530}{408487997986572924289}$$

$$\frac{2470536351695706691150273}{3568842156502352911081681}$$

$$\frac{9147174028201584695660404993}{738256476822867002162862144}$$

$$\frac{404373028971550370031684692 09281}{1981311901313796050853901 1801521}$$

$$\frac{142130019185439765346002547048069394}{144041052884595077187155035625188225}$$

$$\frac{4077551427539061268365818617070082487981}{184324830274606566726090222213840241}$$

$$\frac{29247742836717181569573123126609380958628633}{28854486546227283567381569872895922009146244}$$

$$\frac{1644662183623605030943992459758717959368038089933}{83901072785189703665402414145682292170060509 5881}$$

$$\frac{76795559807444450146033952048248025474377706486132570}{5623685827394653901724022330582419835420495478336929}$$

$$\frac{4207207710529127418283848461340965655463560485366289903505}{603739079570654154039764273913238342923364845621426610 5001}$$

$$\frac{816297679393916005694837838808362431503501229559444925278681793}{148904572022531958307959435081656301977637384156037514895340176}$$

$$\frac{242513738949178952234806483689465816559631390124939658301320990605073}{73775837085713458790733887440677670199204073501807579726644932829689}$$

$$\frac{3730581243032711558064055718833454835524258042187865528369957879076706 2474}{478032325309932556594714214910085243349652938578868570758473383867849 76289}$$

$$\frac{67559659782039617237841184516992302782851604142385500859648938761010393239431661}{14936574510410904335660741815647139057367850205068241677325855756259592064 08025}$$

Figure 4.2. Multiples of $(1, 1)$ on the rational curve $Y^2 = X^3 + X - 1$

Chapter VIII.4), and quite small in general. A technique for finding high rank curves is due to Mestre; its theoretical bases are described in [Mestre, 1986], while the algorithm itself can be found in [Mestre, 1982]. The basic idea is to force the curve to have as many points as possible on its reductions modulo small primes. This implies conditions on the values of the curve coefficients modulo the product of all primes under consideration. While it is conjectured that there are elliptic curves of arbitrarily high rank, the actual record is for a curve whose rank could be proved to be at least 23 ([Martin and McMillen, 1997]).

Altogether these arguments indicate that there are probably not enough points of small height to construct a reasonably large factor base.

Hence the only known attacks on general elliptic curve cryptosystems are the algorithms with square root complexity like those presented in Sections 4.1 and 4.2. However, several approaches have been described which work efficiently for specific classes of curves.

MOV attack

The first such algorithm is due to Menezes, Okamoto and Vanstone, who reduced the elliptic curve logarithm problem to a logarithm problem in a suitable extension of the underlying field. For supersingular curves this reduction results in a subexponential algorithm. Before sketching the method as published in [Menezes et al., 1993a] we restate the problem in the additive formulation for elliptic curves and make an observation to handle the case $p|n$, in which the reduction does not work immediately.

Let E be an elliptic curve defined over the field $\mathbb{F}_q$ of characteristic p, P a point on $E_{\mathbb{F}_q}$ of (not necessarily maximal) order n and $R = lP$ with l to be determined. Similarly to the Pohlig–Hellman method, the logarithm problem can be split into smaller ones if n is composite. Suppose that $n = n_1 n_2$ with n_1 and n_2 coprime. Then it is possible to determine $l \bmod n_1$ and $l \bmod n_2$ separately and to combine the results via the Chinese Remainder Theorem to obtain l:

Consider $n_2 P$, a point of order n_1, and $n_2 R$, and set

$$l_1 = \log_{n_2 P}(n_2 R).$$

Since l is a solution to this logarithm problem it follows that

$$l \equiv l_1 \pmod{n_1}.$$

A similar process determines $l \pmod{n_2}$.

For the reduction to work we need $p \nmid n$; in the light of the remark above, we write $n = p^\nu n'$ with $p \nmid n'$ and have to solve two discrete logarithm problems in groups of order p^ν and n' respectively. In the case p is small (particularly for $p = 2$) the former task can easily be performed by the Pohlig–Hellman method. Then without loss of generality we can assume that n and p are coprime, which implies $E[n] \simeq \mathbb{Z}_n \times \mathbb{Z}_n$ by Proposition 3.37. The MOV reduction works as follows:

1. Determine the smallest integer k such that $E[n] \subseteq E_{\mathbb{F}_{q^k}}$.

2. Compute a point $Q \in E[n]$ such that $\alpha := e_n(P, Q)$ is a primitive n-th root of unity.

3. Set $\beta = e_n(R, Q)$, where e_n is the Weil pairing as described in Section 3.7.

4. $l = \log_P R = \log_\alpha \beta$ in $\mathbb{F}_{q^k}$.

For the correctness of the output note that

$$\beta = e_n(lP, Q) = e_n(P, Q)^l = \alpha^l,$$

and that $\log_\alpha \beta$ is uniquely determined modulo n. The point Q exists by Lemma 3.62. It should be noted that the Weil pairing can be computed in probabilistic polynomial time by an algorithm due to Miller and described in [Menezes, 1993b], pp. 63–68.

In the case of supersingular curves — where $|E_{\mathbb{F}_q}| \equiv 1 \pmod{p}$ by Theorem 3.72 and hence $p \nmid n$ — the authors were able to make Steps 1. and 2. explicit; Table 4.1 summarises the necessary information. The constant c is defined as follows: If $E_{\mathbb{F}_q} \simeq \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2}$ with $n_1|n_2$, then $E_{\mathbb{F}_{q^k}} \simeq \mathbb{Z}_{cn_2} \times \mathbb{Z}_{cn_2}$.

$t = q+1-\lvert E_{\mathbb{F}_q}\rvert$	group structure	n_2	k	c
0	cyclic	$q+1$	2	1
0	$\mathbb{Z}_2 \times \mathbb{Z}_{\frac{q+1}{2}}$	$\frac{q+1}{2}$	2	2
$\pm\sqrt{q}$	cyclic	$q+1\mp\sqrt{q}$	3	$\sqrt{q}\pm 1$
$\pm\sqrt{2q}$	cyclic	$q+1\mp\sqrt{2q}$	4	$q\pm\sqrt{2q}+1$
$\pm\sqrt{3q}$	cyclic	$q+1\mp\sqrt{3q}$	6	$\frac{q+1}{q+1\pm\sqrt{3q}}$
$\pm 2\sqrt{q}$	$\mathbb{Z}_{\sqrt{q}\mp 1} \times \mathbb{Z}_{\sqrt{q}\mp 1}$	$\sqrt{q}\mp 1$	1	1

Table 4.1. Reduction of the discrete logarithm problem for supersingular curves

The explicit knowledge of the group exponent cn_2 of $E_{\mathbb{F}_{q^k}}$ allows a probabilistic polynomial time algorithm for determining Q which results in the following effective reduction:

1. Choose a random point $Q' \in E_{\mathbb{F}_{q^k}}$ and set $Q = \frac{cn_2}{n}Q'$. Compute $\alpha = e_n(P, Q)$.

2. Set $\beta = e_n(R, Q)$.

3. $l' = \log_\alpha \beta$ in F_{q^k}.

4. If $l'P = R$, then $l = l'$. Otherwise α is not a primitive n-th root of unity, go back to 2.

As a consequence supersingular curves with $|t| = 2\sqrt{q}$ or $t = 0$ are no more secure in an ElGamal-like cryptosystem than the underlying field or its quadratic extension and hence do not justify the additional effort of computing in the curve group. Non-supersingular curves with small k should obviously be avoided. A necessary condition on k which is easy to test can be derived from the following proposition.

Proposition 4.2 *Let G be a finite abelian group,*

$$G \simeq \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \text{ with } n_1, n_2 \geq 1,\ n_1|n_2,$$

H a subgroup of G. Then

$$H \simeq \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \text{ with } m_1, m_2 \geq 1,\ m_1|m_2 \text{ and } m_1|n_1,\ m_2|n_2.$$

Warning. The proposition states that if G is a direct inner product of at most two cyclic subgroups G_i with $|G_1|$ a divisor of $|G_2|$, then H is a direct inner product of (at most) the same number of cyclic subgroups H_i with $|H_i|$ a divisor of $|G_i|$. This does *not* mean, however, that H_i is a subgroup of G_i. A simple counter-example is provided by $H = \langle(1,1)\rangle$ which lies "skewly" in $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2}$: $m_1 = 1$, $m_2 = n_2$, but $H \not\subseteq \mathbb{Z}_{n_2}$ for $n_1 > 1$.

Proof of the proposition: The assertion holds in full generality for abelian groups with more than two generators, cf. [Hall Jr., 1959], Theorem 3.3.3. Since most algebra books do not contain the theorem, we give an elementary proof for this special case. Denote by $G[n]$ the n-torsion part of G and by

$$G[n^\infty] := \bigcup_{i=0}^{\infty} G\left[n^i\right]$$

the elements of G which are annulled by a power of n. For a prime divisor p of $|G|$, the subgroup $G[p^\infty]$ is the p-Sylow group of G, i.e. the largest subgroup whose order is a power of p. Recursive applications of Lemma 3.49 yield

$$G \simeq \bigtimes_{p||G|} G[p^\infty] \simeq \bigtimes_{p||G|} (\mathbb{Z}_{p^{\nu_1}} \times \mathbb{Z}_{p^{\nu_2}})$$

where ν_i is the exponent of p in n_i. It is thus sufficient to prove the proposition for the p-Sylow groups; the results can then be recombined via the Chinese Remainder Theorem. Hence, we assume that

$$G \simeq \mathbb{Z}_{p^{\nu_1}} \times \mathbb{Z}_{p^{\nu_2}} \text{ with } 0 \le \nu_1 \le \nu_2.$$

If G is cyclic, i.e. $\nu_1 = 0$, then the assertion follows directly from the fact that each subgroup of a cyclic group is cyclic again and that $|H|||G|$. Otherwise, we still have

$$H \simeq \mathbb{Z}_{p^{\mu_1}} \times \cdots \times \mathbb{Z}_{p^{\mu_s}} \text{ with } 1 \le \mu_1 \le \cdots \le \mu_s$$

by the Fundamental Theorem on Abelian Groups 3.47 and the fact that $|H|||G|$.

We first show that $s \le 2$. Note that

$$G[p] \simeq p^{\nu_1 - 1}\mathbb{Z}_{p^{\nu_1}} \times p^{\nu_2 - 1}\mathbb{Z}_{p^{\nu_2}} \simeq \mathbb{Z}_p \times \mathbb{Z}_p,$$

so $|G[p]| = p^2$. By a similar argument $|H[p]| = p^s$. As obviously $H[p] \le G[p]$, it follows that $s \le 2$.

By possibly admitting the value $\mu_1 = 0$ we can assume that $s = 2$. Then p^{ν_2} is the exponent of G and p^{μ_2} is the exponent of H, so clearly $\mu_2 \le \nu_2$. Suppose that $\mu_1 > \nu_1$ such that $\nu_2 \ge \mu_2 \ge \mu_1 > \nu_1$. Then

$$|G[p^{\mu_1}]| = p^{\nu_1}p^{\mu_1} < p^{\mu_1}p^{\mu_1} = |H[p^{\mu_1}],$$

a contradiction. □

Concerning the necessary condition on k we first note that $E_{\mathbb{F}_{q^k}} \simeq \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2}$ with $n_1 | q^k - 1$ by Rück's Theorem 3.76. Since $E[n] \simeq \mathbb{Z}_n \times \mathbb{Z}_n$ is supposed to be a subgroup of $E_{\mathbb{F}_{q^k}}$, it follows from the proposition that $n | n_1$ and hence $n | q^k - 1$.

A similar reduction, but based on the Tate instead of the Weil pairing is described in [Frey and Rück, 1994]. For this reduction to be into the multiplicative group of the field extension $\mathbb{F}_{q^k}$ it is necessary and sufficient that $n | q^k - 1$.

Hence when devising an elliptic curve cryptosystem with a curve defined over $\mathbb{F}_q$ and based on the discrete logarithm problem in a subgroup of order n, one must make sure that the discrete logarithm problem in $\mathbb{F}_{q^k}^{\times}$ is intractable, where k is the smallest integer such that $n | q^k - 1$.

Trace 1 curves

While the MOV reduction is particularly apt to attack supersingular curves, i.e. curves whose cardinalities are congruent to 1 modulo p by Waterhouse's Theorem 3.72, the curves whose cardinalities are divisible by p allow an even more efficient attack, which was described independently in [Satoh and Araki, 1998], [Semaev, 1998] and [Smart, 1999].

As before, if $n = p^{\nu} n'$ with $p \nmid n'$, then it is sufficient to compute $l = \log_P R$ modulo p^{ν} and modulo n'. Finally to determine l modulo p^{ν} it is sufficient by the Pohlig–Hellman method of Section 4.3 to compute ν discrete logarithms in the subgroup generated by the point $\frac{n}{p} P$ of order p. The new algorithms solve this latter problem in polynomial time by reducing it to a discrete logarithm problem in the *additive* group $\mathbb{F}_p = \{0, \ldots, p-1\}$. Notice that it is only meaningful in the case that p is large, since otherwise discrete logarithms are easy to compute anyway. Specifically, if the curve is defined over the prime field $\mathbb{F}_p$, then $n | p$ is equivalent to $n = p$ by Hasse's Theorem 3.61, which means that the trace of the Frobenius endomorphism is 1.

The presentation in [Semaev, 1998] is the most elementary one and is accessible with the knowledge of Chapter 3.

Xedni calculus

Silverman suggested an algorithm which inverts the index calculus in the sense that first points are lifted to $\mathbb{Q}$ and then a lifted curve is chosen to pass through the points ([Silverman, 1998]). This solves one of the problems mentioned in the paragraphs about the elliptic curve index calculus, since now the lifting procedure is very simple. The basic idea is that a linear dependence over $\mathbb{Z}$ of the lifted points translates into a dependence of the points on the original curve, which can then be used to derive the discrete logarithm. However, the lifted points are likely to be independent, and Silverman suggested to impose inverse Mestre conditions, i.e. to choose a rational curve with as few points as possible on its reductions modulo small primes. This would hopefully increase the likelihood of the curve having smaller rank than expected. Koblitz observed that an efficient xedni calculus could be used to compute discrete logarithms in finite prime fields as well as to factor integers, which would compromise the security of all public key cryptosystems employed today ([Silverman, 1998], Appendix K). A group centred at the University of Waterloo, Canada, is actually investigating the practicality of the algorithm ([Jacobson et al., 1999]).

Certicom ECC challenge

The Canadian company Certicom created an ECC challenge, much in the spirit of the RSA challenge of factoring large integers. The task is to compute discrete logarithms on randomly chosen elliptic curves over prime fields and fields of characteristic 2, and on *Koblitz curves*, which are defined over $\mathbb{F}_2$ and whose points have coordinates in an extension field $\mathbb{F}_{2^m}$ ([Certicom, 1997]). So far, curves over fields with slightly less than 2^{100} elements have been successfully attacked by parallel variants of Pollard's ρ-method (see [Escott, 1998] and [Certicom, 1997]). It was observed in [Gallant et al., 1998] and [Wiener and Zuccherato, 1998] that Koblitz curves have an additional structure which speeds up the ρ-method by a factor of $\sqrt{m}$ compared to a random curve over the field $\mathbb{F}_{2^m}$. The improved algorithms make use of the Frobenius endomorphism $(x, y) \mapsto (x^2, y^2)$.

5 COUNTING POINTS ON ELLIPTIC CURVES

"Trois et deux font cinq. Cinq et sept douze et trois quinze. Bonjour. Quinze et sept vingt-deux. Vingt-deux et six vingt-huit. (...) Vingt-six et cinq trente et un. Ouf! Ça fait donc cinq cent un millions six cent vingt-deux mille sept cent trente et un."
Et j'aime la nuit écouter les étoiles. C'est comme cinq cent millions de grelots...
—Saint-Exupéry

We have seen in the previous chapter that the security of a discrete logarithm based cryptosystem relies mainly on the order of the underlying group, unless special structures allow more efficient algorithms for breaking the system. If the group order is large enough, then square root attacks like Shanks's baby-step giant-step or Pollard's ρ-methods are not applicable. To make the Pohlig–Hellman attack impractical, two different approaches are conceivable.

On one hand, it is possible to choose a group with unknown order, so that the Pohlig–Hellman algorithm does not work. This is a risky game, however, since for no known type of groups there is a theoretical barrier to compute their orders. For instance, the problem is not known to be NP-complete for any class of groups. Hence, there is a certain chance that an adversary already has an algorithm at hands for determining the group order. Moreover, while this attitude allows to encrypt messages, the signature algorithms of Chapter 1 require that the group order be known.

On the other hand, it is a good strategy to make sure that the group order contains a large prime factor to prevent the Pohlig–Hellman attack. In the case of elliptic curves this can be achieved in various ways. First, by the complex multiplication method, curves with suitable orders can be designed specifically ([Atkin and Morain, 1993] and [Lay and Zimmer, 1994]). Second, it is possible to choose special classes of curves whose cardinalities are easy to determine, like supersingular curves (cf. Theorem 3.72), or curves which are defined over a small field, but where the group is chosen over a field extension (cf. Theorem 3.66). While supersingular curves are not recommendable according to Section 4.5, nothing can so far be hold against curves defined over subfields. However, there is a certain reluctance concerning classes of special curves and a widespread belief that the most secure way of selecting a curve is to fix an underlying field, randomly choose a curve, i.e. defining coefficients, and compute the group order until it is divisible by a large prime. This approach is feasible today due to the algorithmic progress made in the past fifteen years.

THROUGHOUT THE CHAPTER WE ASSUME AGAIN THAT $k = \mathbb{F}_q = \mathbb{F}_{p^m}$ IS A FINITE FIELD OF CHARACTERISTIC p AND E AN ELLIPTIC CURVE DEFINED OVER k.

5.1 THE BABY-STEP GIANT-STEP ALGORITHM

We introduced the baby-step giant-step technique in Section 4.1 to compute the discrete logarithm of an element β to the base α in an arbitrary group. In this situation the smallest positive integer k satisfying $\alpha^k = \beta$ has to be determined. Basically the same algorithm can be used to compute the order k of an element α: Here k is the smallest positive integer such that $\alpha^k = 1$. This would allow to determine the order of E_k if the group were cyclic and if we knew a generator of it. The second problem can be solved by testing random elements until a generator is found, the first one by computing the orders of subgroups with two random generators since we already know from Section 3.12 that E_k is generated by two elements. Furthermore, for the cryptographic applications it would be enough to choose a point on E_k and test whether its order has a large prime factor. In any case we need to know how to pick random points on E_k.

Algorithm 5.1 (Picking Points on E_k) *The following probabilistic algorithm picks (almost) uniformly points on E_k. For $p \neq 2$ it needs an expected number of $O(\log q)$, for $p = 2$ an expected number of $O(\log^2 q)$ operations in k.*

1. *Choose a random element $x \in k$; this can be done uniformly.*

2. *Solve the root finding problem $E(x, Y) = 0$ over k. If a solution y exists in k, then $-y - a_1x - a_3$ is the second one. Select randomly one of them to obtain a point $P = (x, y)$ on E_k. Otherwise return to 1.*

The final distribution is not completely uniform since points of order 2 are picked with twice the probability of the other finite points and $\mathcal{O}$ is never chosen. The first problem can be remedied by preliminarily computing the X-coordinates of points of order 2 and selecting them with only half of the probability of the other values in Step 1. The second problem is in fact unimportant since we are only interested in finite points during the following algorithms. In any case the distribution is uniform on all but at most four points, which can be neglected.

Proof: The probability of success in one round of the algorithm is

$$\begin{aligned} & \frac{\text{number of different } X\text{-coordinates of finite points on } E_k}{q} \\ \geq & \frac{(|E_k| - 1)/2}{q} \\ \geq & \frac{q - 2\sqrt{q}}{2q} \quad \text{by Hasse's Theorem} \\ = & \frac{1}{2} - \frac{1}{\sqrt{q}} \\ \geq & c \end{aligned}$$

for any constant $c < \frac{1}{2}$ and sufficiently large q. So the expected number of rounds of the algorithm is in $O(1)$.

It remains to estimate the time needed for one round, i.e. the expected time for finding a root of a quadratic equation $f = Y^2 + aY + b$ in k. We have seen in Proposition 3.67 how the solvability of a quadratic equation can be decided effectively. Suppose that f has its roots in k. Then for $p \neq 2$, the roots can be found by the Berlekamp–Rabin algorithm in expected $O(\log q)$ operations in k (for a description of the algorithm, see e.g. [Menezes, 1993a], pp. 22–23). If $p = 2$ and $a = 0$, then the double root of f is given by $b^{2^{m-1}}$; if $a \neq 0$, the roots of f are distinct and can be computed by the deterministic Berlekamp trace algorithm with $O(\log^2 q)$ operations (see [Menezes, 1993a], pp. 23–24). □

Neglecting for a while the problem that E_k is not cyclic we turn our attention to the order of a single point. Its computation can be simplified considerably if some extra information is known.

Algorithm 5.2 (Order of Point with Known Torsion) *Suppose that G is an additively written group with neutral element O and $P \in G$ a point of n-torsion, where the factorisation of n is known. Then the following algorithm computes the order of P in $O(\log^2 n)$ group operations:*
While $\frac{n}{l}P = O$ for a prime divisor l of n, replace n by $\frac{n}{l}$.

Proof: Obviously this algorithm computes the order of P. If

$$n = \prod_{i=1}^{r} l_i^{\nu_i}$$

is the factorisation of n, then the algorithm requests at most

$$s = \sum_{i=1}^{r} \nu_i$$

computations of multiples of P. Each multiplication of P by the Double and Add Algorithm 1.1 requires $O(\log n)$ group operations. Since $s \leq \log_2 n$, this proves the assertion. □

With fewer preliminary information we can use a baby-step giant-step technique if only some upper bound on the order of P is known. We present a quite general algorithm, which we will need later as a subroutine for computing the exact number of points on E_k.

Algorithm 5.3 (Order of Arbitrary Point) *Let G be a group and $P \in G$. Let further $1 \leq C < B$, $1 \leq L$ and $0 \leq l_1 < L$ be integers. The following algorithm computes the smallest element $l \in [C, B]$ with $lP = \mathcal{O}$ and $l \equiv l_1 \pmod L$ if such an element exists. If furthermore the order of LP is not larger than $\frac{B-C+2}{L} - 2$, then it is output, too. The algorithm needs $O\left(\sqrt{\frac{B-C}{L}}\right)$ group operations and stores $O\left(\sqrt{\frac{B-C}{L}}\right)$ group elements.*

1. *Let*

$$\begin{aligned} C_1 &= \min\{c \in [C, B] : c \equiv l_1 \pmod L\} \quad \textit{and} \\ B_1 &= \max\{c \in [C, B] : c \equiv l_1 \pmod L\}. \end{aligned}$$

   If C_1 and B_1 do not exist, stop. If $C_1 = B_1$ and $C_1 P = \mathcal{O}$, then output $l = C_1$ and stop. If $C_1 = B_1$ and $C_1 P \neq \mathcal{O}$, then stop. Otherwise compute $P_1 = LP$, $s = \left\lceil \sqrt{\frac{B_1 - C_1}{L} + 1} \right\rceil$ and sP_1.

2. *Compute (iP_1, i) for $0 \leq i < s$ and sort them by the first component.*

3. *For all $0 \leq j < s$ compute recursively*

$$-C_1 P - jsP_1 = (-C_1 P - (j-1)sP_1) - sP_1$$

   and look up this group element in the precomputed list of Step 2. The result is a lexicographically ordered list of matches (j, i); possible matches (j, i) with $C_1 + (js + i)L > B_1$ must be discarded.

4. *If no match is found, stop. Otherwise output the first match (j, i) in the form $l = C_1 + (js + i)L$. Moreover if two subsequent matches (j_1, i_1) and (j_2, i_2) are found, then the order of LP is $(j_2 - j_1)s + (i_2 - i_1)$.*

Proof: For the correctness of the algorithm note that we are looking for the smallest element of

$$\begin{aligned} &\{l \in [C, B] : lP = \mathcal{O}, \quad l \equiv l_1 \pmod L\} \\ = \; &\left\{ C_1 + kL : k \in \left[0, \frac{B_1 - C_1}{L}\right], (C_1 + kL)P = C_1 P + kP_1 = \mathcal{O} \right\}. \end{aligned}$$

A match (j,i) means that

$$\begin{aligned} & -C_1P - jsP_1 = iP_1 \\ \Leftrightarrow\quad & C_1P + (js+i)P_1 = \mathcal{O}. \end{aligned}$$

Since any element in the interval $\left[0, \frac{B_1-C_1}{L}\right]$ has a representation $js+i$ with $0 \le i,j < s$ we are sure not to overlook a match.

It is obvious that if we find two matches, then the algorithm determines the order of P_1 correctly. This case occurs exactly if $\operatorname{ord} P_1 \le \frac{B_1-C_1}{L}$. By $C_1 \le C + (L-1)$ and $B_1 \ge B - (L-1)$ we conclude that $\frac{B_1-C_1}{L} \ge \frac{B-C+2}{L} - 2$, so $\operatorname{ord} P_1$ is output whenever it is at most $\frac{B-C+2}{L} - 2$.

The assertions on the time and space complexity are the usual ones as proven in Section 4.1. □

To handle the case of two generators P and P' we need an algorithm to determine the order of P' in the factor group $E_k/\langle P\rangle$. Again a baby-step giant-step approach is successful.

Algorithm 5.4 (Order of Point in Factor Group) *Let G be a group and $H = \langle P\rangle$ a cyclic normal subgroup, P' another point of G. Suppose that $l = \operatorname{ord}_G P$ and $l' = \operatorname{ord}_G P'$ are known. The following algorithm determines $\operatorname{ord}_{G/H} P'$ with $O\left(\sqrt{\tilde{l}}\log^2 \tilde{l}\right)$ group operations and $O\left(\sqrt{l}\right)$ stored group elements, where $\tilde{l} = \max\{l, l'\}$.*

1. *Let $h = l'$ and $s = \left\lceil \sqrt{l} \right\rceil$. Factor l', e.g. by trial division.*

2. *Compute $\mathcal{B} = \{iP : 0 \le i < s\}$ and sort as usual. Let $\mathcal{G} = \{jsP : 0 \le j < s\}$.*

3. *For any prime divisor p' of l' test by Step 4. if $\frac{h}{p'}P' \in \langle P\rangle$. If this is the case, replace h by $\frac{h}{p'}$ and repeat the test, otherwise proceed to the next prime divisor. The final value of h is the desired order.*

4. *For any $Q \in \mathcal{G}$ look up $\frac{h}{p'}P' - Q$ in $\mathcal{B}$. If one look-up is successful, then $\frac{h}{p'} \in \langle P\rangle$, otherwise $\frac{h}{p'} \notin \langle P\rangle$.*

Proof: Note that the algorithm is a sort of combination between Algorithms 5.2 and 5.3. We know that $l'P' = \mathcal{O}$, so $l'P' \in \langle P\rangle$ and for the correctness of the algorithm only Step 4. remains to be checked. To see that it yields the desired result notice that by construction

$$\begin{aligned} \mathcal{B} + \mathcal{G} &= \{(i+js)P : 0 \le i,j < s\} \\ &= \left\{ iP : 0 \le i < \left\lceil \sqrt{l} \right\rceil^2 \right\} \\ &= \langle P\rangle, \end{aligned}$$

and a match in Step 4. is equivalent to $\frac{h}{p'}P' \in \mathcal{B} + \mathcal{G} = \langle P\rangle$.

For the time complexity we see as usual that one execution of Step 4. needs $O(\log l')$ multiplications for computing $\frac{h}{p'}P'$ and $O(\sqrt{l}\log l)$ group operations for the subtractions and the look-ups; the step is called at most $\log_2 l'$ times. Factoring l' by trial division takes $O\left(\sqrt{l'}\right)$ divisions, each of which is at most as costly as a group inversion. Finally the computation and sorting of $\mathcal{B}$ takes $O(\sqrt{l}\log l)$ operations, so that the total complexity is

$$O\left(\log^2 l' + \sqrt{l}\log l \log l' + \sqrt{l'}\right) \subseteq O\left(\sqrt{\tilde{l}}\log^2 \tilde{l}\right).$$

The result on the space complexity is again the usual one. □

All the subroutines being formulated, we can easily state an algorithm for computing $|E_k|$.

Algorithm 5.5 (Cardinality of E_k) *Let E be an elliptic curve defined over $k = \mathbb{F}_q$. The following probabilistic algorithm determines $|E_k|$ with expected $O\left(\sqrt[4]{q}\log^2 q\log\log q\right)$ group operations, storing $O\left(\sqrt[4]{q}\right)$ group elements.*

1. *Select a random point $P \in E_k\backslash\{\mathcal{O}\}$ by Algorithm 5.1. Call Algorithm 5.3 with $C = q + 1 - \lfloor 2\sqrt{q}\rfloor$, $B = q + 1 + \lfloor 2\sqrt{q}\rfloor$, $L = 1$ and $l_1 = 0$. If only one $r \in [C, B]$ with $rP = \mathcal{O}$ is found, then $|E_k| = r$. Otherwise $l = \operatorname{ord}_G P$ is known.*

2. *Select a second random point $P' \in E_k\backslash\{\mathcal{O}\}$ and call Algorithm 5.3 again, this time with $L = l$ and $l_1 = 0$. If only one $r' \in [C, B]$ with $r'P' = \mathcal{O}$ is found, then $|E_k| = r'$. Otherwise the order L' of lP' is known, and the order l' of P' can be computed by Algorithm 5.2 as a divisor of lL'. Possibly exchange the roles of P and P' such that $l \geq l'$.*

3. *By Algorithm 5.4 determine $t = \operatorname{ord}_{E_k/\langle P\rangle} P'$. If $lt > 2\lfloor 2\sqrt{q}\rfloor$, then $|E_k|$ is the unique number in $[C, B]$ which is divisible by lt. Otherwise return to Step 2.*

Proof: First let us verify that if the algorithm terminates, it produces a correct result. Note that $|E_k|$ must be comprised in the interval $[C, B]$ by Hasse's Theorem 3.61 and furthermore, after fixing P, is a multiple of l. In Step 1. we determine whether only one possible candidate satisfying these conditions exists, so if the algorithm stops in this step — which is exactly the case for $l > 2\lfloor 2\sqrt{q}\rfloor$ — it outputs a correct result. Otherwise $l \leq 2\lfloor 2\sqrt{q}\rfloor$ is known. In Step 2. we choose a second point P' and test if there are several numbers in the Hasse interval which are a simultaneous multiple of l and l', which is the order of P'. Since $|E_k|$ must have this property, we are done if only one candidate satisfying these conditions is found. Finally in Step 3. we compute

$$lt = \operatorname{ord}_{E_k} P \operatorname{ord}_{E_k/\langle P\rangle} P' = |\langle P, P'\rangle|,$$

which must divide $|E_k|$.

Note that the algorithm stops when it stumbles on a pair (P, P') of generators of E_k, so there is a non-zero probability of termination.

The running time of one round of the algorithm is in $O(\sqrt[4]{q}\log^2 q)$ as a direct consequence of the complexities given for Algorithms 5.2 to 5.4; notice that the factorisation of lL' needed for Algorithm 5.2 can be obtained by $O(\sqrt{lL'}) \subseteq O(\sqrt[4]{q})$ trial divisions. However, the expected number of rounds is more difficult to compute due to the various possibilities of termination. To ease the argumentation we analyse a slightly simplified algorithm, where the jump "return to Step 2." in Step 3. is replaced by "return to Step 1.". We derive an upper bound for the expected number of rounds from a lower bound on the probability of success within one round. Let $E_k \simeq \mathbb{Z}_m \times \mathbb{Z}_n$ with $m|n$. Then a round in which $l = n$ and $t > \frac{2\lfloor 2\sqrt{q}\rfloor}{n}$ is certainly successful. We determine the probability of choosing a random pair (P, P') with these properties.

1st case: If $n > 2\lfloor 2\sqrt{q}\rfloor$, then any pair with $l = n$ is successful. There are $m\varphi(n)$ points of order n out of $mn - 1$ possible choices for P, so a round is successful with probability at least $\frac{\varphi(n)}{n}$.

2nd case: $n \leq 2\lfloor 2\sqrt{q}\rfloor$
By $|E_k| = mn \geq q + 1 - 2\sqrt{q} = (\sqrt{q} - 1)^2$ we know that $n \geq \sqrt{q} - 1$. So a pair for which $l = n$ and $t \geq 5$ is successful for large q, precisely for

$$5(\sqrt{q} - 1) > 2\lceil 2\sqrt{q}\rceil \Leftrightarrow q > 25.$$

Again the probability for choosing a suitable P is $\frac{\varphi(n)}{n}$. Then consider the number of points P' with $t < 5$:

$$\begin{aligned} |\{P' \in E_k : t < 5\}| &= |\{P' \in E_k : 3P' \in \langle P\rangle \text{ or } 4P' \in \langle P\rangle\}| \\ &\leq |[3]^{-1}(\langle P\rangle)| + |[4]^{-1}(\langle P\rangle)| \\ &= |\langle P\rangle|(|E[3]| + |E[4]|) \\ &\leq n(9 + 16) \quad \text{by Theorem 3.39} \\ &= 25n. \end{aligned}$$

Hence the ratio of suitable P' is at least

$$\frac{mn - 25n}{mn - 1} \geq \frac{(m - 25)n}{mn} = 1 - \frac{25}{m}.$$

Note that

$$q - 2\sqrt{q} < mn \leq 4\sqrt{q}m$$

implies

$$m > \frac{1}{4}\sqrt{q} - \frac{1}{2} \to \infty \quad (q \to \infty),$$

so the ratio of suitable P' is at least

$$1 - \varepsilon$$

for any positive constant ε and sufficiently large q, and the probability of a successful pair (P, P') is asymptotically $\frac{\varphi(n)}{n}$ even in this case.

So the expected number of rounds before a successful stop is at most

$$\sum_{i=1}^{\infty} i\left(1-\frac{\varphi(n)}{n}\right)^{i-1}\frac{\varphi(n)}{n} = \frac{\varphi(n)}{n}\left(\frac{\partial}{\partial c}\frac{1}{1-c}\right)\bigg|_{c=1-\frac{\varphi(n)}{n}}$$
$$= \frac{\varphi(n)}{n}\frac{1}{(1-c)^2}\bigg|_{1-c=\frac{\varphi(n)}{n}} = \frac{n}{\varphi(n)}.$$

It remains to give an upper bound on $\frac{n}{\varphi(n)}$ in terms of q, which is not trivial since

$$\frac{n}{\varphi(n)} = \prod_{r|n,\, r \text{ prime}} \frac{1}{1-\frac{1}{r}}$$

depends on the prime factors of n and not only on its size. We use a result from [Rosser and Schoenfeld, 1962], p. 72:

$$\frac{n}{\varphi(n)} < e^C \log\log n + \frac{3}{\log\log n}$$

where $C = 0.5772...$ is Euler's constant, so

$$\frac{n}{\varphi(n)} \in O(\log\log q).$$

Hence the expected time complexity is in

$$O\left(\sqrt[4]{q}\log^2 q \log\log q\right) \subseteq O\left(\sqrt[4]{q}\log^3 q\right) \subseteq O\left(q^{\frac{1}{4}+\varepsilon}\right)$$

group operations for any $\varepsilon > 0$. □

Remark. It can be shown that in the second case the probability of a random pair (P, P') to generate the whole group is at least

$$\begin{aligned}\frac{\varphi(m)\varphi(n)}{mn} &> \left(e^C \log\log n + \frac{3}{\log\log n}\right)^{-2} \\ &\geq \left(e^C \log\log(4\sqrt{q}) + \frac{3}{\log\log(4\sqrt{q}}\right)^{-2}.\end{aligned}$$

For $q \approx 10^{25}$ this value is about 2.1 %, so it takes on average no more than 50 rounds of the algorithm until a generating pair is found. For cyclic curves, which fall under the first case, only about 8 rounds suffice on average.

Implementing a similar algorithm, Müller succeeded in computing $|E_k|$ for a finite prime field k of size about 10^{25} ([Müller, 1991], p. 103). He observed that a large portion, about 80 %, of all considered curves were cyclic ([Müller, 1991], pp. 107–109), so repeating Step 1. was in fact sufficient to compute their cardinalities.

5.2 SCHOOF'S ALGORITHM

A theoretical breakthrough for the problem of computing the group order of an elliptic curve was achieved by Schoof, who published the first polynomial time algorithm, which is even deterministic ([Schoof, 1985]).. It is based on Hasse's Theorem, or to be more precise its proof.

Recall that if t denotes $(q+1)-|E_k|$ and $\varphi : (x,y) \mapsto (x^q, y^q)$ the Frobenius endomorphism, then t is the unique integer such that

$$\varphi^2 - t\varphi + q = 0; \tag{5.1}$$

moreover,

$$|t| \leq 2\sqrt{q}$$

by Theorem 3.61. It follows from the last inequality that it is sufficient to determine $t \pmod L$, where L is any integer greater than $4\sqrt{q}$. Now the problem can be broken into several smaller ones: Determine $t \pmod l$ with l a prime different from 2 and p, and the product of all considered primes being larger than $4\sqrt{q}$. The partial results can be combined via the Chinese Remainder Theorem to yield the exact value for t.

To compute $t \pmod l$, note that specialising (5.1) to the l-torsion points $E[l]$ yields

$$\varphi_l^2 - t\varphi_l + q = 0, \tag{5.2}$$

where φ_l is the restriction of φ to $E[l]$ and the equality has to be seen as an identity in the $\mathbb{Z}_l$-algebra $\mathrm{End}(E)|_{E[l]}$ (see the proof of Hasse's Theorem 3.61).

Denote by s the residue of q after reducing modulo l; then for $0 \leq \tau < l$ we have to test which of the relations

$$\varphi_l^2 + s = \tau\varphi_l \tag{5.3}$$

holds. If τ_0 is a solution, then $t \equiv \tau_0 \pmod l$.

To test whether a relation of polynomial or rational functions holds on $E[l]$, we use the division polynomial ψ_l.

Lemma 5.6 *Let l be an odd prime not equal to p and $f = u + vY \in k[E]$ with $u, v \in k[X]$. Then the following assertions are equivalent:*

1. *$f(P) = 0$ for all $P \in E[l]$.*

2. *ψ_l divides u and v in $k[X]$.*

Proof: First of all note that $\psi_l \in k[X]$ by Corollary 3.54, so the assertion of the lemma makes sense. By Proposition 3.57,

$$\operatorname{div} \psi_l = \langle E[l] \rangle - l^2 \langle \mathcal{O} \rangle.$$

Thus f is zero on $E[l]$ if and only if $\frac{f}{\psi_l}$ has no finite poles, which by Proposition 2.34 means that $\frac{f}{\psi_l} \in k[E]$. Let $\frac{f}{\psi_l} = a + bY$ with $a, b \in k[X]$. Then $f = u + vY = a\psi_l + b\psi_l Y$, so $u = a\psi_l$ and $v = b\psi_l$ are divisible by ψ_l. □

Thus testing a polynomial identity on $E[l]$ boils down to testing whether the polynomial (in normalised form) is divisible by ψ_l.

If the polynomial has a special form, then it is also easily possible to test whether it vanishes on *any* l-torsion point:

Lemma 5.7 *Let l be an odd prime not equal to p and $f \in k[E]$ with either $f \in k[X]$ or $\frac{f}{\psi_2} \in k[X]$. Write*

$$\tilde{f} = \begin{cases} f & \text{if } f \in k[X] \\ \frac{f}{\psi_2} & \text{otherwise} \end{cases}$$

Then the following assertions are equivalent:

1. *There is a point $P \in E[l]$ such that $f(P) = 0$.*
2. $\gcd\left(\tilde{f}, \psi_l\right) \neq 1$

Proof: Assume first that $f \in k[X]$, and let $P = (x, y) \in E[l]$ such that $f(P) = 0$. It follows that $X - x$ divides f in $k[X]$. On the other hand, $\psi_l(P) = 0$, so $X - x$ divides ψ_l in $k[X]$ as well, and the greatest common divisor is not trivial. The converse direction is proved in an analogous way. Suppose now that $f = \psi_2 \tilde{f}$ with $\tilde{f} \in k[X]$. Note that $\psi_2(P) \neq 0$ for all $P \in E[l]$ since l is odd. Thus $f(P) = 0$ for an l-torsion point P is equivalent to $\tilde{f}(P) = 0$, and the assertion follows from the previous case. □

To test the relation (5.3) we have to add φ^2 and s; so in a first place we have to check if possibly $\varphi^2 = \pm s$. In his original approach, Schoof first examines if there exists an l-torsion point P such that $\varphi^2(P) = \pm sP$ by the test of Lemma 5.7. Otherwise one knows that $\varphi^2 \neq \pm s$. Unfortunately, since Lemma 5.7 is valid only for very special polynomials, this approach requires a case distinction between even and odd characteristic and furthermore the use of the normal forms of Table 2.2. We follow a slightly different approach and test whether $\varphi^2(P) = \pm sP$ holds on *all* l-torsion points P; then we can use Lemma 5.6, which is valid for all polynomials in $k[E]$ and allows a unified presentation for even and odd characteristic.

1st step: Test for $\varphi_l^2 = \pm s$

A necessary condition for $\varphi_l^2 = \pm s$ is $X(\varphi^2) = X(s)$ on $E[l]$, which means that $X^{q^2} - g_s = 0$ on $E[l]$. By Propositions 3.52 and 3.51, 4.,

$$g_s = X - \frac{\psi_{s-1}\psi_{s+1}}{\psi_s^2} \in k(X),$$

so applying Lemma 5.6 we test if

$$\psi_s^2(X^{q^2} - X) + \psi_{s-1}\psi_{s+1} \equiv 0 \pmod{\psi_l}.$$

If the relation does not hold, we know that $\varphi_l^2 \notin \{s, -s\}$, so we can use the generic addition formulae, see the second step. Otherwise, we have to distinguish between two cases: First, there may be a finite point $P \in E[l]$ with $\varphi_l^2(P) = -sP$; second, we may have $\varphi_l^2 = s$. Notice that *a priori* in the first case we do not necessarily have $\varphi_l^2 = -s$, since there might be finite points $P, Q \in E[l]$ with $\varphi_l^2(P) = -sP$ and $\varphi_l^2(Q) = sQ$. Assuming, however, that $\varphi_l^2(P) = -sP$ for a finite l-torsion point P, the specialisation of (5.2) yields $\mathcal{O} = \tau_0\varphi_l(P)$ and hence $\tau_0 = 0$ since $\varphi_l(P) \neq \mathcal{O}$. So we indeed have either $\varphi_l^2 = -s$, which means $\tau_0 = 0$, or $\varphi_l^2 = s$. In the second case, $\tau_0 \neq 0$, and (5.3) implies

$$2s = \tau_0\varphi_l \quad \Leftrightarrow \quad \varphi_l = \frac{2s}{\tau_0}, \tag{5.4}$$

where the inverse of τ_0 has to be seen in $\mathbb{Z}_l$. Substituting this value for φ_l in (5.3) yields

$$\frac{4s^2}{\tau_0^2} + s = 2s \quad \Leftrightarrow \quad \tau_0^2 = 4s.$$

So τ_0 is a square root of $4s$ in $\mathbb{Z}_l$. We first test whether s is a quadratic residue modulo l, i.e. $(\frac{s}{l}) = 1$. If this is not the case, we are sure that $\varphi_l^2 = -s$ and $\tau_0 = 0$. Otherwise let ω be a root of s, which can be determined by simple trial and error because l is small. By (5.4) this case can be decided by checking whether $\varphi_l = \pm\omega$. So we first test if

$$\psi_\omega^2(X^q - X) + \psi_{\omega-1}\psi_{\omega+1} \equiv 0 \pmod{\psi_l}. \tag{5.5}$$

If the relation does not hold, then $\varphi_l^2 = -s$ and $\tau_0 = 0$. Otherwise $\varphi_l^2 = s$, and we have to distinguish between the cases $\varphi_l = \omega$ and $\varphi_l = -\omega$ by looking at the second coordinates. By Proposition 3.55, 2.,

$$h_\omega = Y + \frac{\psi_{\omega+2}\psi_{\omega-1}^2}{\psi_2\psi_\omega^3} + (3X^2 + 2a_2X + a_4 - a_1Y)\frac{\psi_{\omega-1}\psi_{\omega+1}}{\psi_2\psi_\omega^2}.$$

Clearing the denominators, we compute

$$\begin{aligned} & \psi_2\psi_\omega^3(Y(\varphi) - Y(\omega)) \\ = \; & \psi_2\psi_\omega^3(Y^q - h_\omega) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (5.6) \\ = \; & \psi_2\psi_\omega^3(Y^q - Y) - \psi_{\omega+2}\psi_{\omega-1}^2 - (3X^2 + 2a_2X + a_4 - a_1Y)\psi_{\omega-1}\psi_\omega\psi_{\omega+1} \end{aligned}$$

and eliminate higher powers of Y to obtain a polynomial $u + vY$ with $u, v \in k[X]$. If ψ_l divides u and v, we conclude that $\varphi_l = \omega$ and $\tau_0 = 2\omega$, otherwise $\varphi_l = -\omega$ and $\tau_0 = -2\omega$.

Remark. In the case of odd characteristic it is more efficient to choose the normal form with $a_1 = a_3 = 0$ and to represent h_ω by the simpler first formula of Proposition 3.55 as

$$h_\omega = \frac{\psi_{\omega+2}\psi_{\omega-1}^2 - \psi_{\omega-2}\psi_{\omega+1}^2}{2\psi_2\psi_\omega^3}.$$

Writing

$$Y^q = Y(X^3 + a_2X^2 + a_4X + a_6)^{\frac{q-1}{2}}$$

it follows that $v = 0$ by Proposition 3.51, 4., so only one divisibility test has to be performed. I am unaware of a possibility for the general case to avoid the second test.

2nd step:
If the test of the first step has failed, we know that $\varphi_l^2 \neq \pm s$, so $\varphi^2 \neq \pm s$, and we can use the generic addition formulae to compute $\varphi^2 + s$:

$$\begin{aligned}
\alpha &= \psi_2\psi_s^3(Y(\varphi^2) - Y(s)) \\
&= \psi_2\psi_s^3(Y^{q^2} - h_s) \\
&= \psi_2\psi_s^3(Y^{q^2} - Y) - \psi_{s+2}\psi_{s-1}^2 \\
&\quad -(3X^2 + 2a_2X + a_4 - a_1Y)\psi_{s-1}\psi_s\psi_{s+1} \text{ by Proposition 3.55, 2.} \\
\beta &= \psi_2\psi_s^3(X(\varphi^2) - X(s)) \\
&= \psi_2\psi_s^3(X^{q^2} - g_s) \\
&= \psi_2\psi_s^3(X^{q^2} - X) + \psi_2\psi_{s-1}\psi_s\psi_{s+1} \text{ by Proposition 3.52} \\
\lambda &= \frac{\alpha}{\beta} \\
g_\varphi &= \psi_s^2\beta^2X(\varphi^2 + s) \\
&= \psi_s^2\beta^2(-X^{q^2} - g_s + \lambda^2 + a_1\lambda - a_2) \\
&= \psi_s^2\left(\left(\left(-X^{q^2} - X - a_2\right)\beta + a_1\alpha\right)\beta + \alpha^2\right) + \beta^2\psi_{s-1}\psi_{s+1} \\
&\quad \text{by Proposition 3.52} \\
h_\varphi &= \psi_s^2\beta^3Y(\varphi^2 + s) \\
&= -\alpha(g_\varphi - X^{q^2}\psi_s^2\beta^2) - (Y^{q^2} + a_3)\psi_s^2\beta^3 - a_1\beta g_\varphi \\
&= \psi_s^2\left(-\left(Y^{q^2} + a_3\right)\beta + \alpha X^{q^2}\right)\beta^2 - (\alpha + a_1\beta)g_\varphi
\end{aligned}$$

g_φ and h_φ can be computed modulo ψ_l once for the rest of this step. We then test whether $\varphi_l^2 + s = \tau\varphi_l$ for $-\frac{l-1}{2} \leq \tau \leq \frac{l-1}{2}$ until we find the value τ_0. For each positive value of τ, we compute

$$\begin{aligned}
&\psi_s^2\beta^2\psi_\tau^2(X^q, Y^q)\left((X(\varphi^2 + s) - X(\tau\varphi)\right) = \\
&\qquad \psi_\tau^2(X^q, Y^q)g_\varphi - \psi_s^2\beta^2\left(\psi_\tau^2(X^q, Y^q)X^q - \psi_{\tau-1}(X^q, Y^q)\psi_{\tau+1}(X^q, Y^q)\right)
\end{aligned}$$

and test whether it is divisible by ψ_l. Notice that $\gcd(\psi_s^2, \psi_l) = 1$: If $s = 0$, this is trivial, otherwise $\gcd(s, l) = 1$ and $E[s] \cap E[l] = \{\mathcal{O}\}$ contains no finite point. Likewise, $\gcd(\psi_\tau^2(X^q, Y^q), \psi_l) = \gcd(\psi_\tau^{2q}, \psi_l) = 1$. However, β may be zero in a finite l-torsion point, namely for $Q \in E[l]$ with $\varphi^2(Q) = \pm sQ$. But in this second step there is at least one finite point $P \in E[l]$ with $\varphi^2(P) \neq \pm sP$. Hence, if divisibility of the above polynomial by ψ_l holds, then $(\varphi^2+s)(P) = \pm\tau\varphi(P)$ for this point P. Since P also satisfies (5.3), i.e. $(\varphi^2+s)(P) = \tau_0\varphi(P)$ and $\varphi(P)$ is a point of order l, this implies $\tau_0 \equiv \pm\tau \pmod{l}$. To distinguish between $\tau_0 = \tau$ and $\tau_0 = -\tau$ we have to compare the Y-coordinates by computing

$$\begin{aligned}
&\psi_s^2\beta^3\psi_2(X^q,Y^q)\psi_\tau^3(X^q,Y^q)\left(Y(\varphi^2+s)-Y(\tau\varphi)\right) = \\
&\psi_2(X^q,Y^q)\psi_\tau^3(X^q,Y^q)h_\varphi \\
&\quad-\psi_s^2\beta^3\left(\psi_2(X^q,Y^q)\psi_\tau^3(X^q,Y^q)Y^q+\psi_{\tau+2}(X^q,Y^q)\psi_{\tau-1}^2(X^q,Y^q)\right) \\
&\quad-\psi_s^2\beta^3(3X^{2q}+2a_2X^q+a_4-a_1Y^q)\psi_{\tau-1}(X^q,Y^q)\psi_\tau(X^q,Y^q)\psi_{\tau+1}(X^q,Y^q)
\end{aligned}$$

and testing for divisibility by ψ_l. If the test is successful, then $\tau_0 = \tau$, otherwise $\tau_0 = -\tau$.

Remark. Again in the case of odd characteristic, every other divisibility test can be omitted since the polynomials are either in $k[X]$ or in $Yk[X]$. Also in the general case it may be desirable to factor powers of ψ_2 from the division polynomials. A careful arrangement of the computations and storing of intermediate results are important; all intermediate polynomials can be reduced modulo ψ_l. In fact this is the only way of computing X^q, X^{q^2}, Y^q and Y^{q^2} by a repeated square and multiply technique and frequent reductions modulo ψ_l and E.

For a better overview we summarise the algorithm again:

Algorithm 5.8 (Schoof) *The following algorithm computes the order of E_k with $O\left(\log^6 q\right)$ multiplications and inversions in k, storing $O\left(\log^3 q\right)$ field elements.*

1. *Determine a set of prime numbers $\mathcal{L}$, different from 2 and p, such that*

$$\prod_{l\in\mathcal{L}} l > 4\sqrt{q}.$$

Precompute a list of division polynomials ψ_i for $2 \le i \le \max\mathcal{L}$ by the recursion formulae of Proposition 3.53. For all $l \in \mathcal{L}$, determine $\tau_0 \pmod{l}$ by Steps 2. to 5.

2. *Let $s = q \bmod l$. Compute $X^q \bmod \psi_l$, $X^{q^2} \bmod \psi_l$, $Y^q \bmod (E,\psi_l)$ and $Y^{q^2} \bmod (E,\psi_l)$ by a square and multiply algorithm.*

3. *Determine*

$$\psi_s^2(X^{q^2}-X)+\psi_{s-1}\psi_{s+1} \bmod \psi_l.$$

If the result is not zero, go to Step 4. Otherwise compute $\left(\frac{s}{l}\right)$. If $\left(\frac{s}{l}\right) = -1$, then $\tau_0 = 0$. Otherwise determine $\omega \in \mathbb{Z}_l$ such that $\omega^2 = s$ by trial and error. Compute

$$\psi_\omega^2(X^q-X)+\psi_{\omega-1}\psi_{\omega+1} \bmod \psi_l.$$

If this polynomial is not zero, then $\tau_0 = 0$. *Otherwise check whether*

$$\psi_2\psi_\omega^3(Y^q - Y) - \psi_{\omega+2}\psi_{\omega-1}^2 - (3X^2 + 2a_2X + a_4 - a_1Y)\psi_{\omega-1}\psi_\omega\psi_{\omega+1} \mod (E, \psi_l) = 0.$$

If the relation holds, then $\tau_0 = 2\omega$, *otherwise* $\tau_0 = -2\omega$. *If* τ_0 *has been determined in this step, return to Step 2. and continue with the next prime in* $\mathcal{L}$.

4. *Construct a list of* $\psi_i(X^q, Y^q)$ *for* $2 \le i \le \frac{l-1}{2}$ *by the recursion formulae of Proposition 3.53, which are equally valid for* $\psi_i \circ \varphi$ *as for* ψ_i *itself.*

5. *Compute the following polynomials modulo* ψ_l *and possibly* E:

$$\begin{aligned} \alpha &= \psi_2\psi_s^3(Y^{q^2} - Y) - \psi_{s+2}\psi_{s-1}^2 \\ &\quad -(3X^2 + 2a_2X + a_4 - a_1Y)\psi_{s-1}\psi_s\psi_{s+1} \\ \beta &= \psi_2\psi_s^3(X^{q^2} - X) + \psi_2\psi_{s-1}\psi_s\psi_{s+1} \\ g_\varphi &= \psi_s^2\left(\left(\left(-X^{q^2} - X - a_2\right)\beta + a_1\alpha\right)\beta + \alpha^2\right) + \beta^2\psi_{s-1}\psi_{s+1} \\ h_\varphi &= \psi_s^2\left(-\left(Y^{q^2} + a_3\right)\beta + \alpha X^{q^2}\right)\beta^2 - (\alpha + a_1\beta)g_\varphi \end{aligned}$$

Repeat the following computations for $1 \le \tau \le \frac{l-1}{2}$ *until* τ_0 *is found. Compute*

$$\psi_\tau^2(X^q, Y^q)g_\varphi - \psi_s^2\beta^2\left(\psi_\tau^2(X^q, Y^q)X^q - \psi_{\tau-1}(X^q, Y^q)\psi_{\tau+1}(X^q, Y^q)\right) \mod (E, \psi_l).$$

If the result is not zero, then go on with the next value of τ. *Otherwise compute*

$$\begin{aligned} &\psi_s^2\beta^3\psi_2(X^q, Y^q)\psi_\tau^3(X^q, Y^q)\left(Y(\varphi^2 + s) - Y(\tau\varphi)\right) = \\ &\psi_2(X^q, Y^q)\psi_\tau^3(X^q, Y^q)h_\varphi \\ &-\psi_s^2\beta^3\left(\psi_2(X^q, Y^q)\psi_\tau^3(X^q, Y^q)Y^q + \psi_{\tau+2}(X^q, Y^q)\psi_{\tau-1}^2(X^q, Y^q)\right) \\ &-\psi_s^2\beta^3(3X^{2q} + 2a_2X^q + a_4 - a_1Y^q)\psi_{\tau-1}(X^q, Y^q)\psi_\tau(X^q, Y^q)\psi_{\tau+1}(X^q, Y^q) \\ &\mod (E, \psi_l). \end{aligned}$$

If the result is zero, then $\tau_0 = \tau$, *otherwise* $\tau_0 = -\tau$.

6. *By the Chinese Remainder Theorem compute* t *as the unique number in* $[-2\sqrt{q}, 2\sqrt{q}]$ *satisfying* $t \equiv \tau_0 \pmod{l}$ *for all* $l \in \mathcal{L}$. *Then* $|E_k| = q + 1 - t$.

For a complexity analysis we have to know the number of steps needed for elementary polynomial operations. Since additions in k are usually for free (they need $O(\log q)$ bit operations) we only count multiplications and inversions (which need $O(\log^2 q)$ bit operations) as elementary operations in k.

Lemma 5.9 *Let f and g be polynomials of $k[X]$, $d_f = \deg f \geq \deg g = d_g$, $\alpha \in k$. Then the number of elementary operations in k needed to compute*

- *$f + g$ is zero,*
- *αf is in $O(d_f)$,*
- *fg is in $O(d_f d_g)$ and*
- *a polynomial division with remainder of f by g is in $O((d_f - d_g)d_g)$.*

Proof: The assertions for addition and scalar multiplication are trivial. For fg note that $(d_f + 1)(d_g + 1)$ monomials have to be multiplied and added up. The computation of $f = ag + b$ with $\deg b < d_g$ by the usual algorithm is about as costly as the computation of ag, where $\deg a = d_f - d_g$; for details see [Cohen, 1993], p. 110. □

Lemma 5.10 *Let f, g and h be reduced elements of $k[E]/(\psi_l)$, i.e. $f = f_1 + f_2Y$, $g = g_1 + g_2Y$ and $h = h_1 + h_2Y$ with $f_1, f_2, g_1, g_2, h_1, h_2 \in k[X]$ of degree less than $\deg \psi_l$, and let $\alpha \in k$. Then the number of elementary operations to compute reduced representations of*

- *$f + g$ is zero,*
- *αf is in $O(l^2)$ and*
- *fg is in $O(l^4)$*

Proof: Note that the degree of ψ_l is $\frac{l^2-1}{2} \in O(l^2)$. The first two assertions are again trivial. To compute fg, write

$$\begin{aligned} fg &= (f_1 + f_2Y)(g_1 + g_2Y) \\ &= f_1g_1 + f_2g_2(X^3 + a_2X^2 + a_4X + a_6) \\ &\quad +(f_1g_2 + f_2g_1 - f_2g_2(a_1X + a_3))Y. \end{aligned}$$

So we have to compute a fixed number of products of polynomials in one variable of degree $O(l^2)$. By the previous lemma this needs $O(l^4)$ operations, and the two divisions with remainder of the resulting polynomials of degree $O(l^2)$ by ψ_l require another $O(l^4)$ operations by the same lemma. □

Proof of the complexity of Algorithm 5.8: For the first step of Schoof's algorithm it is recommendable to use $\mathcal{L} = \{p_2, \ldots, p_n\}$ as the set of the first primes different from 2 and p. We have to estimate n and p_n (to simplify the presentation we assume that p_i is indeed the i-th prime; thus for small p our estimates for n, e. g., may differ by 1). We must choose n such that

$$\prod_{i=2}^{n} p_i > 4\sqrt{q}$$

or equivalently

$$\vartheta(p_n) := \log\left(\prod_{i=1}^{n} p_i\right) > \log(8\sqrt{q}).$$

By [Rosser and Schoenfeld, 1962], p. 70,

$$\vartheta(p_n) < p_n\left(1 + \frac{1}{2\log p_n}\right) < 2p_n,$$

so it is sufficient to choose

$$p_n \approx \frac{1}{2}\log(8\sqrt{q}) \in O\left(\log q\right).$$

Hence the list of ψ_i can be computed with $O(p_n) = O(\log q)$ multiplications and divisions of polynomials of degree at most $\frac{p_n^2-1}{2} \in O(\log^2 q)$. Thus the total complexity for this step is $O(\log^5 q)$ by Lemma 5.9. For a single prime l, the computation of X^q, X^{q^2}, Y^q and Y^{q^2} modulo ψ_l in Step 2. takes $O(\log q)$ multiplications in $k[E]/(\psi_l)$, which can be performed in $O(l^4 \log q)$. Step 3. requires a constant number of multiplications in $k[E]/(\psi_l)$, the evaluation of the Legendre symbol $(\frac{s}{l})$ and the computation of a square root modulo l by trial and error. The first task requires $O(l^4)$ elementary operations and dominates the last two ones. The list of division polynomials in X^q and Y^q is constructed in Step 4. by $O(l)$ multiplications in $k[E]/(\psi_l)$ and thus needs $O(l^5)$ elementary operations. Then performing Step 5. for a single value of τ takes a constant number of multiplications and needs $O(l^4)$ operations. The whole step is thus finished in time $O(l^5)$. Taking into account that $l \in O(\log q)$ and that Steps 2. to 5. have to be repeated $|\mathcal{L}| \in O(\log q)$ times, the whole algorithm needs $O(\log^6 q)$ operations in k. The time needed for Chinese remaindering is again dominated by the other steps.

We need to store $O(\log q)$ division polynomials and a constant number of other polynomials of degree at most l^2, so the space complexity is in $O(\log^3 q)$ field elements. □

Remark. The complexity of $O(\log^6 q)$ *field* operations amounts to $O(\log^8 q)$ *bit* operations. In the original description by Schoof the time complexity is $O(\log^9 q)$ bit operations because he computes the division polynomials

$$\psi_i(X^q, Y^q) = (\psi_i(X,Y))^q \mod \psi_l$$

independently from one another by squaring and multiplying. This accounts for $O(\log q)$ multiplications in $k[E]/(\psi_l)$ per division polynomial, while the technique described here, which has been suggested independently by several authors, needs $O(1)$ multiplications per division polynomial. The approach of evaluating ψ_i at the previously computed X^q and Y^q mod ψ_l would even raise the complexity to $O(\log^2 q)$ multiplications per division polynomial.

The algorithm has been implemented by Müller for q a large prime. He was able to use all primes up to 13 while the computations with 17 failed ([Müller, 1991], p. 106). This corresponds to a value of q of about

$$\left(\frac{3\cdot 5\cdot 7\cdot 11\cdot 13}{4}\right)^2 \approx 2^{24}.$$

To push the applicability of Schoof's algorithm a little further, Buchmann and Müller suggested an algorithm combining Schoof's and Shanks's techniques. In a first step, t is computed modulo a set of primes $\mathcal{L}$ by Schoof's algorithm, thus yielding that $t \equiv l_1 \pmod{L}$ with $L = \prod_{l\in\mathcal{L}} l$. Then the first step of Shanks's algorithm is performed: Select a random point P, and by Algorithm 5.3 with $C = q+1-2\sqrt{q}$, $B = q+1+2\sqrt{q}$, L and l_1 determine the elements $r \in [C, B]$ with $rP = \mathcal{O}$ and $r \equiv l_1 \pmod{L}$. Since at least one such element $r = |E_k|$ exists, Algorithm 5.3 is sure to find a match. If this match is unique, we know $|E_k|$. Otherwise the second output of the algorithm is $\operatorname{ord}(LP)$, from which we can determine $\operatorname{ord} P$ by Algorithm 5.2 since $\operatorname{ord} P | L \operatorname{ord}(LP)$. This step can be repeated until a point P with $\operatorname{ord} P \geq \sqrt{q}-1$ is found, then Steps 2. and 3. of Shanks's algorithm are executed to compute the order of a second point P'. Unfortunately the information obtained by Schoof's algorithm does not help to do the second phase (Step 4.) of Shanks's algorithm, but in practice this is not a serious restriction since this phase is needed only seldomly. Note that at this stage at most four values are possible for $|E_k|$, namely the multiples of $\operatorname{ord} P \geq \sqrt{q} - 1$ in the interval $[q + 1 - 2\sqrt{q}, q + 1 + 2\sqrt{q}]$ of length $4\sqrt{q}$.

When using a fixed set of primes $\mathcal{L}$ the complexity of the algorithm is clearly the same as that of Shanks's original method. The record obtained with this algorithm was for a value of q of about $5 \cdot 10^{32}$ according to [Müller, 1991], p. 107.

5.3 ELKIES PRIMES

Although Schoof's algorithm is polynomial, the high exponent in its complexity, which is due to the rapid growth of the degrees of the division polynomials, makes it impractical for larger fields. In at first unpublished works, Atkin and Elkies suggested improvements to the algorithm, which we relate in this and the following sections. Recently, Elkies gave an account of his ideas in [Elkies, 1998]. Good further references are [Müller, 1995], [Schoof, 1995] and [Lercier, 1997a], where many of the details we have to omit here can be found.

As before, let E be defined over $k = \mathbb{F}_q$ and $K = \overline{k}$ be the algebraic closure of k. Let l be an odd prime different from p. We recall that the restriction φ_l of the Frobenius endomorphism to $E[l]$ satisfies an equation

$$\varphi_l^2 - \tau_0\varphi_l + s = 0 \tag{5.7}$$

with $s \equiv q \pmod{l}$, and that we need to determine its trace τ_0. The basic idea of the improvements is to work with a divisor of the division polynomial of smaller degree. Since ψ_l is irreducible in $\mathbb{Z}[X, a_1, a_3, a_2, a_4, a_6]$, there are no "generic" factors, and divisors have to be found individually for each curve after substituting the specific values for the a_i. Then, in principle, it is possible to factor the division polynomial, but this again involves costly computations modulo ψ_l.

Recall that if $E[l] = S \dot{\cup} \overline{S} \dot{\cup} \{\mathcal{O}\}$, then

$$\psi_l = \prod_{P \in S} (X - X(P)).$$

We consider non-trivial subgroups C of $E[l]$, which are exactly the cyclic subgroups of order l of E. Such groups are called "l-groups" henceforth, a term usually applied to all groups whose order is a power of l. In a similar way to the division polynomial ψ_l we form the polynomial

$$f_C = \prod_{P \in S_C} (X - X(P)) \in K[X],$$

where $S_C := S \cap C$, so that $C = S_C \dot{\cup} \overline{S_C} \dot{\cup} \{\mathcal{O}\}$. By its definition, f_C is a divisor of ψ_l in $K[X]$, but it is of computational relevance only when it lies in $k[X]$.

Proposition 5.11 *The following assertions are equivalent:*

1. $f_C \in k[X]$
2. $\varphi_l(C) \subseteq C$
3. $\varphi_l(C) = C$, *i.e. C is invariant under the Frobenius endomorphism.*
4. *C is an eigenspace of φ_l with eigenvalue $\alpha \in \mathbb{Z}_l^\times$*

Proof: Assertions 2. and 3. are equivalent by a simple counting argument since φ_l is injective.

Denote by $\varphi_{K/k}$ the Frobenius automorphism of K/k and its canonical extension to $K[X]$, i.e.

$$\varphi_{K/k} : \sum_{i=0}^{n} a_i X^i \mapsto \sum_{i=0}^{n} a_i^q X^i.$$

Then $f_C \in k[X]$ is equivalent to $\varphi_{K/k}(f_C) = f_C$. Since

$$\varphi_{K/k}(f_C) = \prod_{P \in S_C} (X - X(P)^q) = \prod_{P \in S_C} (X - X(\varphi_l(P))),$$

this means that φ_l permutes the X-coordinates of the points in C and hence leaves C invariant. This shows the equivalence of 1. and 3.

Finally the assertions 3. and 4. are equivalent because C is cyclic; then $\varphi_l(C) = C$ means that a generator P of C is mapped by φ_l to another generator, which must be of the form αP with $\alpha \in \mathbb{Z}_l^\times$. Notice that zero can never be an eigenvalue of φ_l because the map is injective. □

Hence there is a divisor $f_C \in k[X]$ of degree $\frac{l-1}{2}$ of ψ_l if an l-group is invariant under the action of the Frobenius endomorphism, or equivalently if φ_l has an eigenvalue $\alpha \in \mathbb{Z}_l$.

Assume for the moment that we know such an f_C. Then we can determine the corresponding eigenvalue by testing which of the relations $\varphi_l = \pm\alpha$ holds on C for $\alpha \in \left\{1, \ldots, \frac{l-1}{2}\right\}$. This is the exact analogue of the first step of Schoof's algorithm (see page 134), except that now all computations can be performed modulo f_C instead of ψ_l. Once we know α, we can easily determine the value of τ_0: If P is the corresponding eigenvector, then (5.7) implies

$$\mathcal{O} = (\varphi_l^2 - \tau_0\varphi_l + s)(P) = (\alpha^2 - \tau_0\alpha + s)P,$$

and thus $\alpha^2 - \tau_0\alpha + s = 0$ in $\mathbb{Z}_l$ because P is of order l. Then

$$\tau_0 = \alpha + s\alpha^{-1},$$

where the inverse of α has to be computed in $\mathbb{Z}_l$.

Two questions have to be answered to complete the algorithm sketched above: First, how can we determine if a prime is of Elkies type, i.e. φ_l has an eigenvalue in $\mathbb{Z}_l$? Of course, this means that the characteristic polynomial $X^2 - \tau_0 X + s$ of φ_l has a zero in $\mathbb{Z}_l$, which amounts to $\tau_0 - 4s$ being zero or a quadratic residue. This should happen for about half of the primes l, but since we do not know τ_0, this is not a useful criterion. Second, how do we determine the polynomial f_C without explicitly factoring ψ_l? These two questions are addressed in Section 5.4. Moreover, we show in Section 5.5 that even if φ_l does not have an eigenvalue in $\mathbb{Z}_l$, we can obtain partial information on τ_0. We then formulate the complete algorithm in Section 5.6 and report on practical experiences.

5.4 ISOGENIES AND MODULAR POLYNOMIALS

In this section we answer the question whether there are l-groups invariant under the Frobenius endomorphism. It turns out that the action of φ_l on an l-group C can be characterised by the j-invariant of an elliptic curve E/C, such that there is an isogeny with kernel C from E to E/C. Finally, the question reduces to determining whether a certain polynomial has a root in $k[X]$. The proofs of most of the results of this section are beyond the scope of this book; this is particularly true for the final computation of the divisor f_C, which requires some knowledge about complex elliptic curves and modular forms.

Our first task is a mere technicality, namely we have to generalise the concepts of rational maps and isogenies introduced in Section 3.1 to rational maps and isogenies between distinct elliptic curves.

Definition 5.12 *Let E and E' be elliptic curves which are defined over k. A* rational map *from E to E' is a pair of rational functions $\alpha \in K(E) \times K(E)$ that satisfies the equation E', i.e. $E' \circ \alpha = 0$. In other words, a rational map* *is a point on the curve $E'_{K(E)}$. It is* defined over k *if it actually lies on $E'_{k(E)}$. An* isogeny *is a rational map which is furthermore a group homomorphism. E and E' are called* isogenous *if there is an isogeny from E to E'; they are called* k-isogenous *if the isogeny is defined over k.*

It is now a simple exercise to prove the results of Sections 3.1 and 3.2 in the setting of general isogenies by replacing E by E', P by P', and so on where appropriate. The notion of "being isogenous" makes sense because it defines an equivalence relation; only the symmetry is not trivial, but it can be shown that if $\alpha : E \to E'$ is an isogeny of degree m, then there is an isogeny $\hat{\alpha} : E' \to E$, the so-called *dual isogeny,* such that $\hat{\alpha} \circ \alpha = [m]$, see [Silverman, 1986], Section III.6.

Since the kernel of the isogeny (α_1, α_2) consists of the poles of α_1 or α_2 (see the remarks after Definition 3.1) and a rational function has only finitely many poles by Corollary 2.30, the kernel of any non-zero isogeny must be finite. Conversely, Vélu showed in [Vélu, 1971] that for any finite subgroup C of E, there is an elliptic curve E' and an isogeny $\alpha : E \to E'$ of kernel C; E' is unique up to isomorphisms. The isogeny is given by

$$\begin{aligned}\alpha_1(P) &= X(P) + \sum_{Q \in C \setminus \{\mathcal{O}\}} (X(P+Q) - X(Q)) \\ \alpha_2(P) &= Y(P) + \sum_{Q \in C \setminus \{\mathcal{O}\}} (Y(P+Q) - Y(Q)),\end{aligned}$$

or with the notation of Section 3.1:

$$\alpha_1 = \sum_{Q \in C} X \circ \tau_Q - c_1 \quad \text{and} \quad \alpha_2 = \sum_{Q \in C} Y \circ \tau_Q - c_2$$

$$\text{with} \quad c_1 = \sum_{Q \in C \setminus \{\mathcal{O}\}} X(Q) \in K \quad \text{and} \quad c_2 = \sum_{Q \in C \setminus \{\mathcal{O}\}} Y(Q) \in K.$$

The latter formulae show that α_1 and α_2 are indeed rational functions and that they have poles of order 2 and 3, respectively, in exactly the points of C because τ_Q is unramified (see Lemmata 3.14 and 3.16). Since the leading coefficients of α_1 and α_2 satisfy $l(\alpha_1) = l(\alpha_2) = 1$, it follows that $\alpha_2^2 - \alpha_1^3$ has a pole of order at most 5 in $\mathcal{O}$ (see the remarks after Definition 3.40). If the order of the pole is exactly 5, then $\alpha_2^2 - \alpha_1^3 - l(\alpha_2^2 - \alpha_1^3)\alpha_1\alpha_2$ has a pole of order at most 4 in $\mathcal{O}$, and continuing to add suitable multiples of α_1^2, α_2, α_1 and 1 we see that there is a rational function

$$E'(\alpha_1, \alpha_2) = \alpha_2^2 + a'_1\alpha_1\alpha_2 + a'_3\alpha_2 - (\alpha_1^3 + a'_2\alpha_1^2 + a'_4\alpha_1 + a'_6)$$

with a zero in $\mathcal{O}$. The invariance of α_1 and α_2 under translations by points in C shows that $E'(\alpha_1, \alpha_2)$ has zeros in all points of C, thus it cannot have any poles. Then by Proposition 2.34 and Corollary 2.35, $E'(\alpha_1, \alpha_2)$ is a constant and must be zero. It can be shown that E' is non-singular, hence it defines an elliptic curve, and from $E' \circ \alpha = 0$ it follows that α is a rational map. Furthermore, α maps $\mathcal{O}$ to $\mathcal{O}$, which implies that it is an isogeny (see [Silverman, 1986], Section III.4).

The isogeny and the coefficients of E' can be determined explicitly using the addition formulae of Table 2.3 to obtain rational expressions for $X \circ \tau_Q$ and $Y \circ \tau_Q$. The computations are tedious, but elementary. We report only the results. Let $C = S \dot{\cup} \overline{S} \dot{\cup} R \dot{\cup} \{\mathcal{O}\}$ with $R = (E[2] \cap C) \backslash \{\mathcal{O}\}$. We use the following notation (see also Section 3.3):

$$\begin{aligned}
DX &= 2Y + a_1 X + a_3 \\
DY &= 3X^2 + 2a_2 X + a_4 - a_1 Y \\
t(Q) &= \begin{cases} DY(Q) & \text{for } Q \in R \\ (2DY + a_1 DX)(Q) & \text{for } Q \in S \cup \overline{S} \end{cases} \\
u(Q) &= (DX)^2(Q) \\
t &= \sum_{Q \in S \cup R} t(Q) \\
u &= \sum_{Q \in S \cup R} (u(Q) + X(Q)t(Q)).
\end{aligned}$$

Since $t(Q) = t(\overline{Q})$ and $u(Q) = u(\overline{Q})$ for $Q \notin E[2]$, these formulae are independent of the choice of S. Then

$$\begin{aligned}
\alpha_1 &= X + \sum_{Q \in S \cup R} \left(\frac{t(Q)}{X - X(Q)} + \frac{u(Q)}{(X - X(Q))^2} \right) \\
\alpha_2 &= Y - \sum_{Q \in S \cup R} \left(u(Q) \frac{DX}{(X - X(Q))^3} + t(Q) \frac{a_1(X - X(Q)) + Y - Y(Q)}{(X - X(Q))^2} \right. \\
&\qquad \left. + \frac{a_1 u(Q) + DX(Q) DY(Q)}{(X - X(Q))^2} \right)
\end{aligned}$$

$$\begin{aligned}
a'_1 &= a_1 \\
a'_3 &= a_3 \\
a'_2 &= a_2 \\
a'_4 &= a_4 - 5t \\
a'_6 &= a_6 - (a_1^2 + 4a_2)t - 7u
\end{aligned} \tag{5.8}$$

The choice of c_1 and c_2 is in fact arbitrary; it has been made to facilitate the derivation of the formulae above.

For an l-group C, denote by E/C an elliptic curve such that there is an isogeny from E to E/C with kernel C; let j/C be the j-invariant of E/C. Since E/C is unique up to isomorphisms, $j/C \in K$ is well-defined (see Table 2.1). The invariance of C under powers of the Frobenius endomorphism is closely related to j/C:

Theorem 5.13 *Let E be a non-supersingular elliptic curve over k which is not k-isogenous to an elliptic curve of j-invariant 0 or 1728, and let C be an l-group. Then the following assertions are equivalent for a positive integer d:*

1. $\varphi_l^d(C) = C$
2. $j/C \in \mathbb{F}_{q^d}$

Proof: Suppose first that φ_l^d constitutes a bijection of C. By the formulae above it is sufficient to show that t, $u \in \mathbb{F}_{q^d}$; then the a_i' and hence j/C are elements of F_{q^d}. So we have to verify that $t^{q^d} = t$ and $u^{q^d} = u$.

$$\begin{aligned} t^{q^d} &= \sum_{Q \in S \cup R} t(Q)^{q^d} = \sum_{Q \in S \cup R} t\left(\varphi_l^d(Q)\right) \\ &\ \text{because the endomorphism } \varphi_l \text{ respects the order of points and} \\ &\ \text{hence the case distinction in the definition of } t \\ &= \sum_{Q \in \varphi_l^d(S) \cup \varphi_l^d(R)} t(Q). \end{aligned}$$

By assumption, φ_l permutes C, so $\varphi_l^d(R) = R$ (in fact, $R = \emptyset$ here because l is odd) and $C = \varphi_l^d(S) \,\dot{\cup}\, \overline{\varphi_l^d(S)} \,\dot{\cup}\, R \,\dot{\cup}\, \{\mathcal{O}\}$, which shows that $t^{q^d} = t$. The proof for u is completely analogous.

To show the converse direction we would need to know more about the endomorphism rings of elliptic curves than can be treated in this book; see [Müller, 1995], Satz 3.10. □

It follows that the existence of an l-group invariant under the Frobenius endomorphism is equivalent to the polynomial $\prod_{C\ l\text{-group}}(X - j/C)$ having a root in k. This polynomial can be determined without explicitly computing the j/C.

Theorem and Definition 5.14 *There is a polynomial $\Phi_l \in \mathbb{Z}[X, Y]$ with the following property: If E is a non-supersingular elliptic curve over k with j-invariant different from 0 and 1728, then*

$$\Phi_l(X, j(E)) = \prod_{C\ l\text{-group}} (X - j/C).$$

Moreover, $\Phi_l(X, j(E))$ is of degree $l + 1$ and has distinct roots. Φ_l is called the l-th modular polynomial.

Proof: See [Müller, 1995], Satz 4.13 and Lemma 4.14; the proof given there for characteristic $p > 3$ carries over to $p = 2$ and $p = 3$. Similarly to the division polynomials, the factorisation property is first proved for elliptic curves over the complex numbers, and then it is shown that the assertions still hold after reduction modulo p. For the assertion on the degree, note that there are $l + 1$ distinct l-groups: There are $l^2 - 1 = (l-1)(l+1)$ points of order l, i.e. generators of l-groups, and each l-group has $l - 1$ generators. □

Corollary 5.15 *If E is a non-supersingular elliptic curve over k which is not k-isogenous to an elliptic curve of j-invariant 0 or 1728, then the number of l-groups invariant under φ_l^d is given by*

$$\deg\left(\gcd\left(X^{q^d} - X, \Phi_l(X, j(E))\right)\right).$$

Proof: By Theorem 5.13, the number of l-groups invariant under φ_l^d equals the number of j/C which are elements of $\mathbb{F}_{q^d}$. But this is exactly the number of (distinct) roots of $\Phi_l(X, j(E))$ in $\mathbb{F}_{q^d}$ by Theorem 5.14. The result now follows from the well-known relation

$$X^{q^d} - X = \sum_{x \in \mathbb{F}_{q^d}} (X - x).$$

□

This solves the first open question of the previous section. The actual computation of the modular polynomials is rather time consuming; it involves handling the Fourier series expansions of certain modular forms, see [Müller, 1995], Chapters 4 and 5. However, it is sufficient to compute them once for all and to store them on disk. In practice, the modular polynomials have two drawbacks: First, their degree in Y is relatively high (actually, they are symmetric in X and Y), so that there are many coefficients to store; and second, these coefficients soon become very large. For example, here is the third modular polynomial, taken from [Schoof, 1995], p. 236:

$$\begin{aligned}\Phi_3(X,Y) \quad = \quad & X^4 + Y^4 - X^3Y^3 + 2232(X^3Y^2 + X^2Y^3)\\ & -1069956(X^3Y + XY^3) + 36864000(X^3 + Y^3)\\ & +2587918086X^2Y^2 + 8900222976000(X^2Y + XY^2)\\ & +452984832000000(X^2 + Y^2) - 770845966336000000XY\\ & +1855425871872000000000(X + Y)\end{aligned}$$

The problem can be solved using equivalent polynomials, which have the same splitting type after substituting $j(E)$ for Y; for example, it is possible to choose the following polynomials for $l = 3$ and $l = 13$ (courtesy of Markus Maurer and Volker Müller):

$$\begin{aligned}G_3(X,Y) \quad = \quad & X^4 + 36X^3 + 270X^2 + (-Y + 756)X + 729\\ G_{13}(X,Y) \quad = \quad & X^{14} + 26X^{13} + 325X^{12} + 2548X^{11} + 13832X^{10} + 54340X^9\\ & +157118X^8 + 333580X^7 + 509366X^6 + 534820X^5\\ & +354536X^4 + 124852X^3 + 15145X^2 + (-Y + 746)X + 13\end{aligned}$$

Once it is known that a prime l is of Elkies type, i.e. there is an l-group C invariant under φ_l, the desired factor f_C of the l-th division polynomial can be derived from the explicit form of the isogeny $\alpha : E \to E/C$. Since the denominator of α_1 has double poles in exactly the points of C and a line of the form $X - X(P)$ has the divisor $\langle P\rangle + \langle\overline{P}\rangle - 2\langle\mathcal{O}\rangle$ (see the example after Definition 2.27), it follows that up to a constant factor, the denominator of α_1 is $\prod_{P\in S}(X - X(P))^2 = f_C^2$ with $C = S\dot{\cup}\overline{S}\dot{\cup}\{\mathcal{O}\}$.

There are different methods for computing isogenies; we briefly mention them here and refer the reader to the literature. An excellent overview is given in [Lercier, 1997a], Chapters 4 to 8; the complexity assertions are taken from [Lercier and Morain, 1996], p. 13. Note that in principle the question is answered by Vélu's formulae (5.7); however, we need algorithms which do not require the explicit knowledge of C. Elkies's first approach is based on computing isogenies over the complex numbers and generalising the resulting algebraic formulae to finite fields by reduction modulo p ([Müller, 1995], Chapters 6 and 7). Its complexity is $O(l^2)$. The use of the normal form $Y^2 = X^3+a_4X+a_6$ restricts the validity of this procedure to $p \notin \{2,3\}$; moreover, since the denominators of occurring rational numbers may be divisible by small primes, we must have $p > l$. So the technique is especially suited for large prime fields. Couveignes described an algorithm based on the so-called "formal group" associated to an elliptic curve ([Couveignes, 1994]); it works in any characteristic, but has complexity $O(l^3)$. Subsequently, Lercier devised a technique for $p = 2$ with the same asymptotic running time, but which is faster in practice ([Lercier, 1996]). Moreover, it is conceptually simpler. Finally Couveignes found an algorithm of complexity $O(l^{2+\varepsilon})$ for any $\varepsilon > 0$, which is also applicable in any characteristic ([Couveignes, 1996]).

5.5 ATKIN PRIMES

Even if no l-group is invariant under φ_l, it is possible to obtain some information on τ_0. Proposition 5.11 shows that this case occurs when $X^2 - \tau_0X + s$, the characteristic polynomial of φ_l, has no root in $\mathbb{Z}_l$, that is $\tau_0^2 - 4s$ is a quadratic non-residue modulo l. Then $X^2 - \tau_0X + s$ has two distinct roots α and α^l in $\mathbb{F}_{l^2}$. We are now interested in the smallest power of φ_l which leaves an l-group invariant or equivalently has an eigenvalue in $\mathbb{Z}_l$. By the considerations of Section 3.9, the characteristic polynomial of φ_l^e is given by $(X - \alpha^e)(X - \alpha^{le})$. The roots of this polynomial are elements of $\mathbb{Z}_l$ for $\alpha^e = (\alpha^e)^l$, i.e. $(\alpha^{l-1})^e = 1$. So the smallest power of φ_l which leaves an l-group invariant is given by the order d of α^{l-1} in $\mathbb{F}_{l^2}^\times$. Furthermore, since φ_l^d has the double eigenvalue $\alpha^d = \alpha^{ld}$, but the (formal) eigenvalues α and α^l of φ_l are different, it follows that the Jordan normal form of φ_l^d is a diagonal matrix, and *all* l-groups are invariant under φ_l^d. Under the general assumptions of Corollary 5.15, the integer d can be determined as the minimal number such that $\Phi_l(X, j(E))$ divides $X^{q^d} - X$. Letting $\alpha^{l-1} = \zeta_d$, a primitive d-th root of unity in $\mathbb{F}_{l^2}$, there are $\varphi(d)$ possibilities for ζ_d, where now φ denotes the Euler function. Since $(\alpha^{l-1})^{l+1} = \alpha^{l^2-1} = 1$, we have that d divides $l+1$, and $\varphi(d) \leq \varphi(l+1) \leq \frac{l+1}{2}$. The candidates for ζ_d can be determined explicitly if a primitive element g of $\mathbb{F}_{l^2}$ is known, that is $\mathbb{F}_{l^2}^\times = \langle g\rangle$. Then the primitive d-th roots of unity are given by $g^{\frac{l^2-1}{d}i}$ with $\gcd(d, i) = 1$. While in general it is a difficult problem to determine primitive elements for finite fields, the task can be performed easily by trial and error in our case since l is small. Also, the ζ_d can be computed once for all and stored on disk. Comparing coefficients in the equation

$$X^2 - \tau_0 X + s = (X - \alpha)(X - \alpha^l)$$

yields the relations

$$\begin{aligned} \tau_0 &= \alpha + \alpha^l = \alpha(1 + \zeta_d), \quad \text{hence} \\ \tau_0^2 &= \alpha^2(1 + \zeta_d)^2, \quad \text{and} \\ s &= \alpha^{l+1} = \alpha^2 \zeta_d. \end{aligned}$$

This implies

$$\tau_0^2 = \frac{s}{\zeta_d}(1 + \zeta_d)^2 = s\left(\frac{1}{\zeta_d} + 2 + \zeta_d\right).$$

Thus ζ_d and ζ_d^{-1} result in the same value of τ_0^2, so there are $\frac{\varphi(d)}{2}$ possibilities for τ_0^2 and $\varphi(d)$ possibilities for τ_0.

5.6 THE SCHOOF–ELKIES–ATKIN ALGORITHM

We now collect the fruits of Sections 5.3 to 5.5 and present the complete algorithm based on the preparations made therein. In a first step we have to test the assumptions of Corollary 5.15.

Test for supersingularity

By Theorem 3.71, an elliptic curve E in characteristic 2 or 3 is supersingular if and only if $j(E) = 0$. In all other cases, Theorem 3.72 leaves at most five possible values for $|E_k|$. We can then pick some random points on E_k by Algorithm 5.1 and test whether their multiples by one of the candidates for $|E_k|$ are $\mathcal{O}$. If this is not the case, then E cannot be supersingular. Otherwise there is a good chance that E is indeed supersingular, and in the light of the results of Section 4.5, it is advisable not to use the curve for cryptographic purposes.

Test for isogenies with $j \in \{0, 1728\}$

Suppose that E is k-isogenous to an elliptic curve E'. Then the explicit construction of E' from the isogeny α sketched in Section 5.4 shows recursively that a_i' is the leading coefficient of a polynomial in $k[\alpha_1, \alpha_2, a_1', \ldots, a_{i-1}']$ for $i = 1, 2, 3, 4, 6$, so that E' is defined over k. Suppose that E' has j-invariant 0 or 1728. The following theorem shows how we can determine a finite (and actually small) number of candidates for $|E_k|$:

Theorem 5.16 *Let E and E' be k-isogenous curves and both of them defined over k. Then $|E_k| = |E_k'|$.*

Proof: Denote by $\alpha = (\alpha_1, \alpha_2)$ the k-isogeny from E to E' and by φ and φ' the Frobenius endomorphisms of E and E', respectively. Since the coefficients of α_1 are elements of k, we have $\alpha_1(X,Y)^q = \alpha_1(X^q, Y^q)$. The same holds for α_2. Hence,

$$\varphi' \circ \alpha = (\alpha_1(X,Y)^q, \alpha_2(X,Y)^q) = (\alpha_1(X^q,Y^q), \alpha_2(X^q,Y^q)) = \alpha \circ \varphi.$$

If $\varphi^2 - t\varphi + q = 0$ is the characteristic equation of φ, this implies

$$0 = \alpha \circ (\varphi^2 - t\varphi + q) = ({\varphi'}^2 - t\varphi' + q) \circ \alpha,$$

and so ${\varphi'}^2 - t\varphi' + q = 0$ because α is surjective by Proposition 3.3. Then $|E_k| = |E'_k|$ by Theorem 3.61. In fact, the converse is also true, see [Tate, 1966], Theorem 1(c). □

Suppose first that $p = 2$ or $p = 3$. Then $j(E') = 0$, so E' is supersingular by Theorem 3.71. Corollary 3.73, together with $|E_k| = |E'_k|$, implies that E must be supersingular as well, which has been excluded by the previous test. If $p > 3$, it is possible to determine the cardinalities of the curves with j-invariant equal to 0 or 1728, which amounts to finding elements of norm q in the imaginary quadratic number fields $Q(i)$ and $Q(\sqrt{-3})$ (see [Müller, 1995], Chapter 9.2). Again we can test if the corresponding multiples of some random points are $\mathcal{O}$. If this does not hold, we are sure that E is not k-isogenous to a curve with j-invariant 0 or 1728. Otherwise there are algorithms to verify the supposed group order (see [Müller, 1995], Chapter 11). Note, however, that for the security of the cryptographic schemes described in Chapter 1 we need not know the group order, but only the order of a cyclic subgroup, usually generated by a random point; the test above checks exactly the orders of random points, and if a point is found whose order is divisible by a large prime, then it can be used for implementing the cryptographic schemes, no matter what the cardinality of the full elliptic curve group is.

Computing t mod l

If the two test above have been negative, we can compute $\tau_0 \equiv t \pmod{l}$ for odd primes l as described in Sections 5.4 and 5.5. In the case of Elkies primes we get the exact value of τ_0, while in the case of Atkin primes we get a list of at most $\frac{l+1}{2}$ possible values. The complexity for an Elkies prime drops from $O(\log^5 q)$ field operations in Schoof's original algorithm (see the complexity analysis of Algorithm 5.8) to $O(\log^3 q)$ field operations, because the computations are performed modulo a polynomial f_C of degree in $O(l)$ instead of $O(l^2)$. Notice that determining f_C by isogeny computations is dominated by $O(l^3)$; see the discussion in Section 5.4.

Refinements include working with prime powers (see [Müller, 1995], Chapter 8, and [Couveignes and Morain, 1994] or [Couveignes et al., 1996]) and powers of 2 (see [Couveignes and Morain, 1994], Section 5, for $p \neq 2$ and [Menezes et al., 1993b] for $p = 2$).

Combining the partial information

Let $\mathcal{L}_1$ denote the set of Elkies primes (or prime powers) and $\mathcal{L}_2$ the set of Atkin primes considered in the previous step, and let $L_1 = \prod_{l \in \mathcal{L}_1} l$, $L_2 = \prod_{l \in \mathcal{L}_2} l$ and $L = L_1 L_2$. Then by the Chinese Remainder Theorem the value of t is uniquely determined modulo L_1; modulo L_2, however, there are many possibilities. Precisely, if ν_l is the number of candidates for $t \bmod l$, there are $\prod_{l \in \mathcal{L}_2} \nu_l$ possible values for $t \bmod L_2$ and finally for $t \bmod L$, and this is usually exponential in $|\mathcal{L}_2|$.

It is now necessary to test which of the possible values for $|E_k|$ is correct; by Theorem 3.61 we can restrict ourselves to representatives in the interval $[q+1-2\sqrt{q}, q+1+2\sqrt{q}]$. Basically the test is the same as above: We choose random points on E_k, multiply them by the supposed cardinality and check whether the result is $\mathcal{O}$. If there is only one number which passes the test, we are done; otherwise it is possible to extend the sets $\mathcal{L}_1$ and $\mathcal{L}_2$. In practice, if L is chosen bigger than $4\sqrt{q}$, the first candidate to pass the test has always been reported to be the correct cardinality. This part of the algorithm can be made faster by scanning the candidates for $|E_k|$ using a baby-step giant-step technique due to Atkin; see [Müller, 1995], Chapter 10.2, and [Lercier, 1997a], Chapter 11.2.

It is remarkable that the inclusion of Atkin primes turns Schoof's polynomial algorithm into an exponential one, at least if $|\mathcal{L}_2|$ is linear in $\log q$ — which is heuristically the case if $\mathcal{L}_1 \cup \mathcal{L}_2$ is chosen as the set of the first primes. However, for instances of practical interest, the exponential algorithm behaves much better.

Implementations

I am aware of two publicly described implementations of the Schoof–Elkies–Atkin algorithm; certainly the commercial developers of elliptic curve cryptosystems dispose of their own codes.

Lehmann, Maurer, Müller and Shoup succeeded in computing the cardinality of an elliptic curve over a prime field with a 425-digit characteristic ([Müller, 1995]; see also [Lehmann, 1994], [Maurer, 1994] and [Lehmann et al., 1994]). They used the prime numbers up to 983 and had to test $5 \cdot 10^6$ possibilities for the cardinality. The running time was over 3000 hours.

Lercier presents an implementation for arbitrary characteristic in [Lercier, 1997a]. His record is for a curve over $\mathbb{F}_{2^{1301}}$, which took a bit over 1200 hours, using prime numbers up to 673. It should be noted that the flow of the algorithm can be quite complex. After a prime power l^k has been investigated, there are in principle three possibilities:

- One can continue with l^{k+1},
- proceed to the next prime or
- stop this phase and try to combine the results obtained so far.

Lercier suggests a dynamic strategy for deciding between the first two possibilities, namely he assigns to each possible computation a number which estimates its cost and performs cheaper tasks first.

Records are a possibility for testing the limits of existing implementations. But the enormous running times of several months — which can be shortened by the expensive use of parallel hardware — make clear that elliptic curves over fields of such a large size are not suited for everyday cryptography. So it seems in order to give some timings for fields of cryptographically important size. A field suggested for use in cryptosystems is $\mathbb{F}_{2^{155}}$; Lercier reports running times of about 90 s for this field ([Lercier, 1997a], Table 12.3). This sounds reasonable, but notice that the computations have to be repeated several times until a cryptographically suitable curve is found. Indeed, to circumvent the Pohlig–Hellman method described in Section 4.3, $|E_k|$ should possibly be prime; then all known attacks at discrete logarithms are exponential, and this provides sufficient security for a field of size about 2^{155}. Depending on the characteristic it may be unavoidable to have some small prime factors in the cardinality of the curve. If $p = 2$, for instance, Propositions 3.41 and 3.38 show that a non-supersingular curve contains a point of order 4. If this point is defined over k, then $|E_k|$ must be divisible by 4. Lercier tested random curves of this type — with the equation $Y^2 + XY = X^3 + a_6$, where a point of order 4 is given by $(\sqrt[4]{a_6}, \sqrt{a_6}) = \left(a_6^{2^{153}}, a_6^{2^{154}}\right)$ — for their cryptographic suitability. He applied an early abort strategy, dropping a curve as soon as the computations modulo an odd prime l or $l = 8$ revealed that l divided $|E_k|$. He found that only five out of 1000 curves were suitable, and he needed about 250 min in total ([Lercier, 1997a], Table 12.6; see also [Lercier, 1997b]).

It can be concluded that finding cryptographically suitable random elliptic curves is still rather costly; but the algorithmic progress of the last years makes this way of choosing curves possible in practice and provides an attractive alternative to the use of special classes of elliptic curves.

References

Und geschicht nichts newes vnter der Sonnen.
Geschicht auch etwas dauon man sagen möcht
Sihe das ist new.
Denn es ist vor auch geschehen in vorigen zeiten
die vor vns gewesen sind.

—Luther (The Bible)

Adleman, L. M. (1994). The function field sieve. In *[Adleman and Huang, 1994]*, pages 108–121.

Adleman, L. M., DeMarrais, J., and Huang, M.-D. (1994). A subexponential algorithm for discrete logarithms over the rational subgroup of the Jacobians of large genus hyperelliptic curves over finite fields. In *[Adleman and Huang, 1994]*, pages 28–40.

Adleman, L. M. and Huang, M.-D., editors (1994). *Algorithmic Number Theory*, volume 877 of *Lecture Notes in Computer Science*, Berlin. Springer-Verlag.

ANSI (1998a). Agreement of symmetric keys on using Diffie–Hellman and MQV algorithms. Working Draft American National Standard: Public Key Cryptography for the Financial Services Industry X9.42-1998, American National Standards Institute. Available at http://grouper.ieee.org/groups/-1363/private/x9-42-10-02-98.zip.

ANSI (1998b). The elliptic curve digital signature algorithm (ECDSA). Working Draft American National Standard: Public Key Cryptography for the Financial Services Industry X9.62-1998, American National Standards Institute. Available at http://grouper.ieee.org/groups/1363/private/x9-62-09-20-98.zip.

ANSI (1999). Key agreement and key transport using elliptic curve cryptography. Working Draft American National Standard: Public Key Cryptography for the Financial Services Industry X9.63-199x, American National Standards Institute. Available at http://grouper.ieee.org/groups/1363/private/-x9-63-01-08-99.zip.

Atkin, A. O. L. and Morain, F. (1993). Elliptic curves and primality proving. *Mathematics of Computation*, 61(203):29–68.

Bézout, E. (1779). *Théorie Générale des Équations Algébriques.* Paris.

Blake, I. F., Fuji-Hara, R., Mullin, R. C., and Vanstone, S. A. (1984). Computing logarithms in finite fields of characteristic two. *SIAM J. Alg. Disc. Meth.*, 5(2):276–285.

Brent, R. P. (1980). An improved Monte Carlo factorization algorithm. *BIT*, 20:176–184.

Brickell, E. F., editor (1993). *Advances in Cryptology — CRYPTO '92*, volume 740 of *Lecture Notes in Computer Science*, Berlin. Springer-Verlag.

Buell, D. A. and Teitelbaum, J. T., editors (1998). *Computational Perspectives on Number Theory: Proceedings of a Conference in Honor of A.O.L. Atkin*, volume 7 of *Studies in Advanced Mathematics.* American Mathematical Society.

Certicom (1997). ECC challenge. http://www.certicom.com/chal/index.htm.

Charlap, L. S. and Robbins, D. P. (1988). An elementary introduction to elliptic curves. CRD Expository Report 31, Institute for Defense Analyses, Princeton.

Cohen, H. (1993). *A Course in Computational Algebraic Number Theory.* Graduate Texts in Mathematics. Springer-Verlag, New York.

Cohen, H., editor (1996). *Algorithmic Number Theory — ANTS-II*, volume 1122 of *Lecture Notes in Computer Science*, Berlin. Springer-Verlag.

Commission of the European Communities (1998). Proposal for a European parliament and council directive on a common framework for electronic signatures. Technical report, European Union. Available in all languages of the European Union; in english at http://europa.eu.int/comm/dg15/en/media/-infso/com297en.pdf.

Coppersmith, D. (1984). Fast evaluation of logarithms in fields of characteristic two. *IEEE Transactions on Information Theory*, 30(4):587–594.

Couveignes, J.-M. (1994). *Quelques calculs en théorie des nombres.* PhD thesis, Université de Bordeaux I. Available at http://www.ufr-mi.u-bordeaux.fr/-~couveign/Publi/Cou94-4.ps.

Couveignes, J.-M. (1996). Computing l-isogenies with the p-torsion. Preprint; available at http://www.ufr-mi.u-bordeaux.fr/~couveign/Publi/Cants96.ps.

Couveignes, J.-M., Dewaghe, L., and Morain, F. (1996). Isogeny cycles and the Schoof–Elkies–Atkin algorithm. Technical Report LIX/RR/96/03, Laboratoire d'Informatique de l'Ecole Polytechnique (LIX), Palaiseau. Available at ftp://lix.polytechnique.fr/pub/submissions/morain/Preprints/isog-cycles.ps.Z.

Couveignes, J.-M. and Morain, F. (1994). Schoof's algorithm and isogeny cycles. In *[Adleman and Huang, 1994]*, pages 43–58.

Deuring, M. (1941). Die Typen der Multiplikatorenringe elliptischer Funktionenkörper. *Abhandlungen aus dem mathematischen Seminar der hamburgischen Universität*, 14:197–272.

Diffie, W. and Hellman, M. E. (1976). New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6):644–655.

ElGamal, T. (1985). A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory*, 31(4): 469–472.

Elkies, N. D. (1998). Elliptic and modular curves over finite fields and related computational issues. In *[Buell and Teitelbaum, 1998]*, pages 21–76.

Enge, A. (1998). Computing discrete logarithms in high-genus hyperelliptic Jacobians in provably subexponential time. Submitted to *Mathematics of Computation.*

Escott, A. (1998). Implementing a parallel Pollard rho attack on ECC. Transparencies of the presentation given at the 2nd Workshop on Elliptic Curve Cryptography at the University of Waterloo; available at http://cacr.math.-uwaterloo.ca/escott.ps.zip.

Frey, G. and Rück, H.-G. (1994). A remark concerning m-divisibility and the discrete logarithm in the divisor class group of curves. *Mathematics of Computation*, 62(206):865–874.

Fulton, W. (1969). *Algebraic Curves.* Mathematics Lecture Note Series. The Benjamin/Cummings Publishing Company, Reading (Massachusetts).

Fumy, W., editor (1997). *Advances in Cryptology — EUROCRYPT '97*, volume 1233 of *Lecture Notes in Computer Science*, Berlin. Springer-Verlag.

Gallant, R., Lambert, R., and Vanstone, S. (1998). Improving the parallelized Pollard lambda search on binary anomalous curves. Preprint.

Gauß, C. F. (1801). *Disquisitiones Arithmeticae.* Gerh. Fleischer Jun., Leipzig.

Gillings, R. J. (1972). *Mathematics in the Time of the Pharaohs.* MIT Press, Cambridge (Massachusetts).

Gordon, D. M. (1993). Discrete logarithms in $GF(p)$ using the number field sieve. *SIAM Journal on Discrete Mathematics*, 6(1):124–138.

Gordon, D. M. and McCurley, K. S. (1993). Massively parallel computation of discrete logarithms. In *[Brickell, 1993]*, pages 312–323.

Hall Jr., M. (1959). *The Theory of Groups.* Macmillan, New York.

Hasse, H. (1934). Abstrakte Begründung der komplexen Multiplikation und Riemannsche Vermutung in Funktionenkörpern. *Abhandlungen aus dem mathematischen Seminar der hamburgischen Universität*, 10:325–348.

Husemöller, D. (1987). *Elliptic Curves.* Graduate Texts in Mathematics. Springer-Verlag, New York.

IEEE (1998). Standard specifications for public key cryptography. Technical Report P1363/D8 (Draft Version 8), Institute of Electrical and Electronics Engineering. Available at http://grouper.ieee.org/groups/1363/index.html.

Jacobson, M. J., Koblitz, N., Silverman, J. H., Stein, A., and Teske, E. (1999). Analysis of the xedni calculus attack. Preprint.

Johnson, D. S., Nishizeki, T., Nozaki, A., and Wolf, H. S., editors (1987). *Discrete Algorithms and Complexity, Proceedings of the Japan–US Joint Seminar, June 4–6, 1986, Kyoto, Japan*, volume 15 of *Perspectives in Computing*, Orlando. Academic Press.

Knuth, D. E. (1981). *The Art of Computer Programming*, volume 2 - Seminumerical Algorithms. Addison-Wesley, Reading (Massachusetts), 2nd edition.

Koblitz, N. (1987). Elliptic curve cryptosystems. *Mathematics of Computation*, 48(177):203–209.

Koblitz, N. (1991). Constructing elliptic curve cryptosystems in characteristic 2. In *[Menezes and Vanstone, 1991]*, pages 156–167.

Koblitz, N. (1993). *Introduction to Elliptic Curves and Modular Forms.* Graduate Texts in Mathematics. Springer-Verlag, New York, 2nd edition.

Koblitz, N. (1994). *A Course in Number Theory and Cryptography.* Graduate Texts in Mathematics. Springer-Verlag, New York, 2nd edition.

Koblitz, N. (1998). *Algebraic Aspects of Cryptography*, volume 3 of *Algorithms and Computations in Mathematics.* Springer-Verlag, Berlin.

Lang, S. (1978). *Elliptic Curves: Diophantine Analysis*, volume 231 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, Berlin.

Lang, S. (1987). *Elliptic Functions.* Graduate Texts in Mathematics. Springer-Verlag, New York, 2nd edition.

Lay, G.-J. and Zimmer, H. G. (1994). Constructing elliptic curves with given group order over large finite fields. In *[Adleman and Huang, 1994]*, pages 250–263.

Lehmann, F., Maurer, M., Müller, V., and Shoup, V. (1994). Counting the number of points on elliptic curves over finite fields of characteristic greater than three. In *[Adleman and Huang, 1994]*, pages 60–70.

Lehmann, F. J. (1994). Implementierung von Algorithmen zur Berechnung modularer Polynome und deren Anwendung im Algorithmus von Atkin. Master's thesis, Universität des Saarlandes, Saarbrücken. Available at ftp://-ftp.informatik.tu-darmstadt.de/pub/TI/reports/lehmann.diplom.ps.gz.

Lercier, R. (1996). Computing isogenies in $GF(2^n)$. In *[Cohen, 1996]*, pages 197–212.

Lercier, R. (1997a). *Algorithmique des courbes elliptiques dans les corps finis.* PhD thesis, École polytechnique, Palaiseau. Available at ftp://lix.polytechnique.fr/pub/lercier/papers/these.ps.Z.

Lercier, R. (1997b). Finding good random elliptic curves for cryptosystems defined over $\mathbb{F}_{2^n}$. In *[Fumy, 1997]*, pages 379–392.

Lercier, R. and Morain, F. (1996). Algorithms for computing isogenies between elliptic curves. To appear in Computational Perspectives on Number Theory, 1997; available at ftp://lix.polytechnique.fr/pub/submissions/morain/-Preprints/isogenies.ps.Z and ftp://lix.polytechnique.fr/pub/lercier/papers/-isogenies.ps.Z.

Lewis, D., editor (1971). *Proceedings of Symposia in Pure Mathematics*, volume 10, Providence (Rhode Island). American Mathematical Society.

Lovorn Bender, R. (1999). Rigorous, subexponential algorithms for discrete logarithms in $GF(p^2)$. To appear in *SIAM J. Discrete Math.*

Martin, R. and McMillen, W. (1997). An elliptic curve over $\mathbb{Q}$ with rank at least 23. Posting to the Number Theory List, see http://listserv.nodak.edu/-archives/nmbrthry.html.

Maurer, M. (1994). Eine Implementierung des Algorithmus von Atkin zur Berechnung der Punktanzahl elliptischer Kurven über endlichen Primkörpern der Charakteristik größer drei. Master's thesis, Universität des Saarlandes, Saarbrücken. Available at ftp://ftp.informatik.tu-darmstadt.de/pub/-TI/reports/maurer.diplom.ps.gz.

Maurer, U. M. and Wolf, S. (1996). On the complexity of breaking the Diffie-Hellman protocol. Technical Report 244, Institute for Theoretical Computer Science, ETH Zürich. Available at ftp://ftp.inf.ethz.ch/pub/publications/-papers/ti/isc/Diffie_Hellman_DL_TR.ps.gz.

McCurley, K. S. (1989). Cryptographic key distribution and computation in class groups. In *[Mollin, 1989]*, pages 459–479.

Menezes, A., editor (1993a). *Applications of Finite Fields*. Kluwer Academic Publishers, Boston/Dordrecht/London.

Menezes, A. (1993b). *Elliptic Curve Public Key Cryptosystems*. Kluwer Academic Publishers, Boston/Dordrecht/London.

Menezes, A. and Vanstone, S. (1990). Isomorphism classes of elliptic curves over finite fields of characteristic 2. *Utilitas Mathematica*, 38:135–153.

Menezes, A. J., Okamoto, T., and Vanstone, S. A. (1993a). Reducing elliptic curve logarithms to logarithms in a finite field. *IEEE Transactions on Information Theory*, 39(5):1639–1646.

Menezes, A. J., Oorschot, P., and Vanstone, S. A. (1997). *Handbook of Applied Cryptography*. CRC Press, Boca Raton.

Menezes, A. J. and Vanstone, S. A., editors (1991). *Advances in Cryptology — CRYPTO '90*, volume 537 of *Lecture Notes in Computer Science*, Berlin. Springer-Verlag.

Menezes, A. J., Vanstone, S. A., and Zuccherato, R. J. (1993b). Counting points on elliptic curves over F_{2^m}. *Mathematics of Computation*, 60(201):407–420.

Mestre, J.-F. (1982). Construction d'une courbe elliptique de rang $\geq$ 12. *Comptes Rendus des Séances de l'Académie des Sciences de Paris, Série I*, 295: 643–644.

Mestre, J.-F. (1986). Formules explicites et minorations de conducteurs de variétés algébriques. *Compositio Mathematica*, 58:209–232.

Miller, V. S. (1986). Use of elliptic curves in cryptography. In *[Williams, 1986]*, pages 417–426.

Mollin, R. A., editor (1989). *Number Theory and Applications*, volume 265 of *NATO ASI Series C: Mathematical and Physical Sciences*, Dordrecht. Kluwer Academic Publishers.

Morain, F. (1997). Classes d'isomorphismes des courbes elliptiques supersingulières en caractéristique $\geq$ 3. *Utilitas Mathematica*, 52:241–253.

Müller, V. (1991). Die Berechnung der Punktanzahl von elliptischen Kurven über endlichen Primkörpern. Master's thesis, Universität des Saarlandes, Saarbrücken. Available at ftp://ftp.informatik.tu-darmstadt.de/pub/TI/reports/vmueller.diplom.ps.gz.

Müller, V. (1995). *Ein Algorithmus zur Bestimmung der Punktanzahl elliptischer Kurven über endlichen Körpern der Charakteristik größer drei*. PhD thesis, Universität des Saarlandes, Saarbrücken. Available at ftp://ftp.informatik.tu-darmstadt.de/pub/TI/reports/vmueller.diss.ps.gz.

Müller, V., Stein, A., and Thiel, C. (1997). Computing discrete logarithms in real quadratic congruence function fields of large genus. To appear in Mathematics of Computation; available at http://www.informatik.tu-darmstadt.-de/TI/Mitarbeiter/vmueller/ffdl.ps.gz.

NIST (1994). Digital signature standard (DSS). Federal Information Processing Standard Publication 186, National Institute of Standards and Technology. Available at http://csrc.nist.gov/fips/fips186.ps.

NIST (1995). Secure hash standard. Federal Information Processing Standard Publication 180-1, National Institute of Standards and Technology. Available at http://csrc.nist.gov/fips/fip180-1.ps.

NIST (1998). Digital signature standard (DSS). Federal Information Processing Standard Publication 186-1, National Institute of Standards and Technology. Available at http://csrc.nist.gov/fips/fips1861.pdf.

Ohta, K. and Pei, D., editors (1998). *Advances in Cryptology — ASIACRYPT '98*, volume 1514 of *Lecture Notes in Computer Science*, Berlin. Springer-Verlag.

Oorschot, P. and Wiener, M. J. (1999). Parallel collision search with cryptanalytic applications. *Journal of Cryptology*, 12(1):1–28.

Pohlig, S. C. and Hellman, M. E. (1978). An improved algorithm for computing logarithms over $GF(p)$ and its cryptographic significance. *IEEE Transactions on Information Theory*, 24(1):106–110.

Pollard, J. M. (1978). Monte Carlo methods for index computation $(\bmod p)$. *Mathematics of Computation*, 32(143):918–924.

Pomerance, C. (1987). Fast, rigorous factorization and discrete logarithm algorithms. In *[Johnson et al., 1987]*, pages 119–143.

Rosser, J. B. and Schoenfeld, L. (1962). Approximate formulas for some functions of prime numbers. *Illinois Journal of Mathematics*, 6:64–94.

Rück, H.-G. (1987). A note on elliptic curves over finite fields. *Mathematics of Computation*, 49(179):301–304.

Satoh, T. and Araki, K. (1998). Fermat quotients and the polynomial time discrete log algorithm for anomalous elliptic curves. *Commentarii Mathematici Universitatis Sancti Pauli*, 47(1):81–92.

Schirokauer, O. (1993). Discrete logarithms and local units. *Philosophical Transactions Royal Society London A*, 345:409–423.

Schnorr, C. P. and Lenstra Jr., H. W. (1984). A Monte Carlo factoring algorithm with linear storage. *Mathematics of Computation*, 43(167):289–311.

Schoof, R. (1985). Elliptic curves over finite fields and the computation of square roots mod p. *Mathematics of Computation*, 44(170):483–494.

Schoof, R. (1987). Nonsingular plane cubic curves over finite fields. *Journal of Combinatorial Theory*, A 46:183–211.

Schoof, R. (1995). Counting points on elliptic curves over finite fields. *Journal de Théorie des Nombres de Bordeaux*, 7:219–254.

Semaev, I. A. (1998). Evaluation of discrete logarithms in a group of p-torsion points of an elliptic curve in characteristic p. *Mathematics of Computation*, 67(221):353–356.

Shafarevich, J. R. (1974). *Basic Algebraic Geometry.* Die Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin.

Shanks, D. (1971). Class number, a theory of factorization and genera. In *[Lewis, 1971]*, pages 415–440.

Shoup, V. (1997). Lower bounds for discrete logarithms and related problems. In *[Fumy, 1997]*, pages 256–266.

Silverman, J. H. (1986). *The Arithmetic of Elliptic Curves.* Graduate Texts in Mathematics. Springer-Verlag, New York.

Silverman, J. H. (1994). *Advanced Topics in the Arithmetic of Elliptic Curves.* Graduate Texts in Mathematics. Springer-Verlag, New York.

Silverman, J. H. (1998). The xedni calculus and the elliptic curve discrete logarithm problem. Preprint; available at http://www.math.brown.edu/˜jhs/-Preprints/XedniCalculus.ps.gz.

Silverman, J. H. and Suzuki, J. (1998). Elliptic curve discrete logarithms and the index calculus. In *[Ohta and Pei, 1998]*, pages 110–125.

Smart, N. P. (1999). The discrete logarithm problem on elliptic curves of trace one. To appear in Journal of Cryptology.

Stinson, D. R. (1995). *Cryptography — Theory and Practice.* Discrete Mathematics and its Applications. CRC Press, Boca Raton.

Tate, J. (1966). Endomorphisms of abelian varieties over finite fields. *Inventiones mathematicae*, 2:134–144.

Teske, E. (1998). Better random walks for Pollard's rho method. Technical Report CORR98-52, Centre for Applied Cryptographic Research, University of Waterloo. Available at http://cacr.math.uwaterloo.ca/techreports/1998/-corr98-52.ps.

UNCITRAL (1998a). Draft uniform rules on electronic signatures. Technical Report A/CN.9/WG.IV/WP.79, United Nations Commission on International Trade Law. Available at http://www.un.or.at/uncitral/english/-sessions/wg_ec/wp-79.htm.

UNCITRAL (1998b). Electronic signatures. Technical Report A/CN.9/WG.-IV/WP.80, United Nations Commission on International Trade Law. Available at http://www.un.or.at/uncitral/english/sessions/wg_ec/wp-80.htm.

Vélu, J. (1971). Isogénies entre courbes elliptiques. *Comptes Rendus de l'Académie des Sciences de Paris, Série A*, 273:238–241.

Waterhouse, W. C. (1969). Abelian varieties over finite fields. *Annales Scientifiques de l'École Normale Supérieure*, 4^e *Série*, 2:521–560.

Weber, D. (1996). Computing discrete logarithms with the general number field sieve. In *[Cohen, 1996]*, pages 391–403.

Wiener, M. J. and Zuccherato, R. J. (1998). Faster attacks on elliptic curve cryptosystems. In *Proceedings of SAC, Workshop on Selected Areas in Cryptography, Lecture Notes in Computer Science.*

Williams, H. C., editor (1986). *Advances in Cryptology — CRYPTO '85*, volume 218 of *Lecture Notes in Computer Science*, Berlin. Springer-Verlag.

Index

> *τοιγὰρ ἐγώ τοι, ξεῖνε πάτερ, δόμον, ὅν με κελεύεις, δείξω. (...)*
> *ἀλλ' ἴθι σιγῇ τοῖον, ἐγὼ δ' ὁδὸν ἡγεμονεύσω.*
>
> —ΟΜΗΡΟΥ

Since their invention in the late seventies, public key cryptosystems have become an indispensable asset in establishing private and secure electronic communication, and this need, given the tremendous growth of the Internet, is likely to continue growing. Elliptic curve cryptosystems represent the state of the art for such systems.

ELLIPTIC CURVES AND THEIR APPLICATIONS TO CRYPTOGRAPHY: An Introduction provides a comprehensive and self-contained introduction to elliptic curves and how they are employed to construct secure public key cryptosystems. Even though the elegant mathematical theory underlying cryptosystems is considerably more involved than for other systems, this text requires the reader to have only an elementary knowledge of basic algebra. The text nevertheless leads to problems at the forefront of current research, featuring chapters on point counting algorithms and security issues. The adopted unifying approach treats with equal care elliptic curves over fields of even characteristic, which are especially suited for hardware implementations, and curves over fields of odd characteristic, which have traditionally received more attention.

ELLIPTIC CURVES AND THEIR APPLICATIONS TO CRYPTOGRAPHY: An Introduction has been used successfully for teaching advanced undergraduate courses. It will be of greatest interest to mathematicians, computer scientists, and engineers who are curious about elliptic curve cryptography in practice, without losing the beauty of the underlying mathematics.

ISBN 0-7923-8589-6

9 780792 385899

0–7923–8589–6