Graduate Texts in Mathematics

Dale Husemöller

Elliptic Curves

椭园曲线［英］

Springer-Verlag
World Publishing Corp

Dale Husemöller

Elliptic Curve

With an Appendix by Ruth Lawrenc

With 44 Illustrations

Springer-Verlag
World Publishing Corp

Dale Husemöller
Department of Mathematics
Haverford College
Haverford, PA 19041
U.S.A.

AMS Classifications: 14-01, 14H25, 14K07, 14L15

Library of Congress Cataloging in Publication Data
Husemöller, Dale.
Elliptic curves.
(Graduate texts in mathematics; 111)
Bibliography: p.
Includes index.
1. Curves, Elliptic. 2. Curves, Algebraic. 3. Group schemes (Mathematics)
I. Title. II. Series.
QA567.H897 1986 512'.33 86-11832

Reprinted by World Publishing Corporation, Beijing, 1990
for distribution and sale in The People's Republic of China only

ISBN 7-5062-0734-6

ISBN 0-387-96371-5 Springer-Verlag New York Berlin Heidelberg
ISBN 3-540-96371-5 Springer-Verlag Berlin Heidelberg New York

To
Robert and Roger
with whom I first learned
the meaning of collaboration

Preface

The book divides naturally into several parts according to the level of the material, the background required of the reader, and the style of presentation with respect to details of proofs. For example, the first part, to Chapter 6, is undergraduate in level, the second part requires a background in Galois theory and the third some complex analysis, while the last parts, from Chapter 12 on, are mostly at graduate level. A general outline of much of the material can be found in Tate's colloquium lectures reproduced as an article in *Inventiones* [1974].

The first part grew out of Tate's 1961 Haverford Philips Lectures as an attempt to write something for publication closely related to the original Tate notes which were more or less taken from the tape recording of the lectures themselves. This includes parts of the Introduction and the first six chapters. The aim of this part is to prove, by elementary methods, the Mordell theorem on the finite generation of the rational points on elliptic curves defined over the rational numbers.

In 1970 Tate returned to Haverford to give again, in revised form, the original lectures of 1961 and to extend the material so that it would be suitable for publication. This led to a broader plan for the book.

The second part, consisting of Chapters 7 and 8, recasts the arguments used in the proof of the Mordell theorem into the context of Galois cohomology and descent theory. The background material in Galois theory that is required is surveyed at the beginning of Chapter 7 for the convenience of the reader.

The third part, consisting of Chapters 9, 10, and 11, is on analytic theory. A background in complex analysis is assumed and in Chapter 10 elementary results on p-adic fields, some of which were introduced in Chapter 5, are used in our discussion of Tate's theory of p-adic theta functions. This section is based on Tate's 1972 Haverford Philips Lectures.

The fourth part, namely Chapters 12, 13, and 14, covers that part of algebraic theory which uses algebraic geometry seriously. This is the theory of endomorphisms and elliptic curves over finite and local fields. While earlier chapters treated an elliptic curve as a curve defined by a cubic equation, here the theory of endomorphisms requires a more subtle approach with varieties and, for some questions of bad reduction, schemes. This part is very carefully covered in the book by Silverman [1985], and thus we frequently do not give detailed arguments. We recommend this book as a reference while reading this part.

The fifth part, consisting of Chapters 15, 16, and 17, surveys recent results in the arithmetic theory of elliptic curves. Here again few proofs are given, but various elementary background results are included for the beginner reader in order to make the main references more accessible. The three chapters include part of Serre's theory of Galois representations including a result of Falting's which played an important role in the proof of the Mordell conjecture, L-functions of elliptic curves over a number field, the special case of complex multiplication, modular curves, and finally the Birch and Swinnerton-Dyer conjecture. There is a progress discussion on the Birch and Swinnerton-Dyer conjecture describing the contributions of Coates and Wiles, of Greenberg, and of Gross and Zagier. We also mention the work of Goldfeld which reduced the effective lower bound question of Gauss for the class number of imaginary quadratic fields to a special case of the conjectural framework of Birch and Swinnerton-Dyer contained in the work of Gross and Zagier.

Finally the book concludes with an appendix by Ruth Lawrence. She did all the hundred or so exercises in the book, and from this extensive work the idea of an appendix evolved. It consists of comments on all the exercises including complete solutions for a representative number. Usually there are just answers or hints on how to proceed together with remarks on the level of difficulty. This appendix should be a great help for the reader starting the subject and wishing to do some of the exercises.

Acknowledgments

Being an amateur in the field of elliptic curves, I would have never completed a project like this without the professional and moral support of a great number of persons and institutions over the long period during which this book was being written.

John Tate's treatment of an advanced subject, the arithmetic of elliptic curves, in an undergraduate context has been an inspiration for me during the last 25 years while at Haverford. The general outline of the project, together with many of the details of the exposition, owe so much to Tate's generous help.

The E.N.S. course by J.-P. Serre of four lectures in June 1970 together with two Haverford lectures on elliptic curves were very important in the early development of the manuscript. I wish to thank him also for many stimulating discussions. Elliptic curves were in the air during the summer seasons at the I.H.E.S. around the early 1970s. I wish to thank P. Deligne, N. Katz, S. Lichtenbaum, and B. Mazur for many helpful conversations during that period. It was the Haverford College Faculty Research Fund that supported many times my stays at the I.H.E.S.

During the year 1974–5, the summer of 1976, the year 1981–2, and the spring of 1986, I was a guest of the Bonn Mathematics Department SFB and later the Max Planck Institute. I wish to thank Professor F. Hirzebruch for making possible time to work in a stimulating atmosphere and for his encouragement in this work. An early version of the first half of the book was the result of a Bonn lecture series on Elliptische Kurven. During these periods, I profited frequently from discussions with G. Harder and A. Ogg.

Conversations with B. Gross were especially important for realizing the final form of the manuscript during the early 1980s. I am very thankful for his encouragement and help. In the spring of 1983 some of the early chapters of the book were used by K. Rubin in the Princeton Junior Seminar, and I thank him for several useful suggestions. During the same time, J. Coates invited me to an Oberwolfach conference on elliptic curves where the final form of the manuscript evolved.

During the final stages of the manuscript, both R. Greenberg and R. Rosen read through the later chapters, and I am grateful for their comments. I would like to thank P. Landweber for a very careful reading of the manuscript and many useful comments.

Ruth Lawrence read the early chapters along with working the exercises. Her contribution was very great with her appendix on the exercises and suggested improvements in the text. I wish to thank her for this very special addition to the book.

Free time from teaching at Haverford College during the year 1985–6 was made possible by a grant from the Vaughn Foundation. I wish to express my gratitude to Mr. James Vaughn for this support, for this project as well as others, during this difficult last period of the preparation of the manuscript.

Contents

Introduction to Rational Points on Plane Curves

This introduction is designed to bring up some of the main issues of the book in an informal way so that the reader with only a minimal background in mathematics can get an idea of the character and direction of the subject.

An elliptic curve, viewed as a plane curve, is given by a nonsingular cubic equation. We wish to point out what is special about the class of elliptic curves among all plane curves from the point of view of arithmetic. In the process the geometry of the curve also enters the picture.

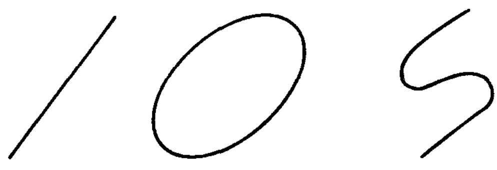
For the first considerations our plane curves are defined by a polynomial equation in two variables $f(x, y) = 0$ with rational coefficients. The main invariant of this f is its degree, a natural number. In terms of plane analytic geometry there is a locus C_f of this equation in the x, y-plane where the definition is that the point (x, y) is on the locus C_f provided it satisfies the equation $f(x, y) = 0$ as real numbers. To emphasize that the locus consists of points with real coordinates (so is in $\mathbf{R}^2$), we denote this real locus by $C_f(\mathbf{R})$ and consider $C_f(\mathbf{R}) \subset \mathbf{R}^2$.

Since some curves C_f, like for example $f(x, y) = x^2 + y^2 + 1$, have an empty real locus $C_f(\mathbf{R})$, it is always useful to work also with the complex locus $C_f(\mathbf{C})$ contained in $\mathbf{C}^2$ even though it cannot be completely pictured geometrically. This is especially true for geometric considerations involving the curve.

For arithmetic the locus of special interest is the set $C_f(\mathbf{Q})$ of rational points $(x, y) \in \mathbf{Q}^2$ satisfying $f(x, y) = 0$, that is, points whose coordinates are rational numbers. The fundamental problem is the description of this set $C_f(\mathbf{Q})$. An elementary question is whether or not $C_f(\mathbf{Q})$ is finite or even empty.

The problem is attacked by a combination of geometric and arithmetic arguments using the inclusions $C_f(\mathbf{Q}) \subset C_f(\mathbf{R}) \subset C_f(\mathbf{C})$. A locus $C_f(\mathbf{Q})$ is either compared with another locus $C_g(\mathbf{Q})$, which is better understood, as we illustrate for lines where $\deg(f) = 1$ and conics where $\deg(f) = 2$ or by internal operations which is the case for elliptic curves.

In terms of the real locus, curves of degree 1, degree 2, and degree 3 can be pictured respectively as follows.



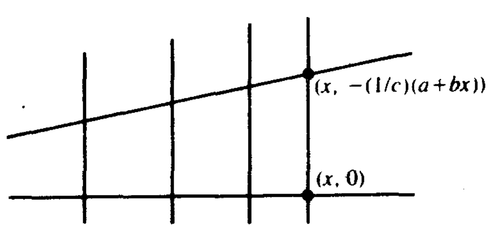
§1. Rational Lines in the Projective Plane

Plane curves C_f can be defined for any nonconstant complex polynomial $f(x, y) \in \mathbf{C}[x, y]$ by the equation $f(x, y) = 0$. For a nonzero constant k the equations $f(x, y) = 0$ and $kf(x, y) = 0$ have the same solutions and define the same plane curve $C_f = C_{kf}$. When f has complex coefficients, there is only a complex locus defined. If f has real coefficients or if f differs from a real polynomial by a nonzero constant, then there is also a real locus with $C_f(\mathbf{R}) \subset C_f(\mathbf{C})$. Such curves are called real curves.

(1.1) Definition. A rational plane curve is one of the form C_f where $f(x, y)$ is a polynomial with rational coefficients.

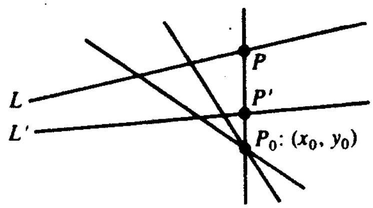
In the case of a rational plane curve C_f we have rational, real, and complex points $C_f(\mathbf{Q}) \subset C_f(\mathbf{R}) \subset C_f(\mathbf{C})$ or loci.

A polynomial of degree 1 has the form $f(x, y) = a + bx + cy$. Assume the coefficients are rational numbers and we begin by describing the set $C_f(\mathbf{Q})$. For c nonzero we can set up a bijective correspondence between rational points on the line C_f and on the x-axis using intersections with vertical lines.



The rational point $(x, 0)$ on the x-axis corresponds to the rational point $(x, -(1/c)(a + bx))$ on C_f. When b is nonzero, the line $C_f(\mathbf{Q})$ can be put in bijective correspondence with the rational points on the y-axis using intersections with horizontal lines. Observe that the vertical or horizontal lines relating rational points are themselves rational lines.

Instead of using parallel lines to relate points on two lines $L = C_f$ and $L' = C_{f'}$, we can use a point $P_0 = (x_0, y_0)$ not on either L or L' and relate points using the family of all lines through P_0. The pair P on L and P' on L' correspond when P, P', and P_0 are all on a line.



If L and L' are rational lines, and if P_0 is a rational point, then for two corresponding points P on L and P' on L' the point P is rational if and only if P' is rational, and this defines a bijection between $C_f(\mathbf{Q})$ and $C_{f'}(\mathbf{Q})$.

Observe that there are special cases of lines through P_0, i.e., those parallel to L or L', which as matters stand do not give a corresponding pair of points between L and L'. This is related to the fact that the two types of correspondence with parallel lines and lines through a point are really the same when viewed in terms of the projective plane, for parallel lines intersect at a point on the "line at infinity." The projective plane is the ordinary Cartesian or affine plane together with an additional line called the line at infinity.

(1.2) Definition. The projective plane $\mathbf{P}_2$ is the set of all triples $w{:}x{:}y$, where w, x, and y are not all zero and the points $w{:}x{:}y = w'{:}x'{:}y'$ provided there is a nonzero constant k with

$$w' = kw, \qquad x' = kx, \qquad y' = ky.$$

As with the affine plane and plane curves we have three basic cases

$$\mathbf{P}_2(\mathbf{Q}) \subset \mathbf{P}_2(\mathbf{R}) \subset \mathbf{P}_2(\mathbf{C})$$

consisting of triples proportional to $w{:}x{:}y$, where $x, y, w \in \mathbf{Q}$ for $\mathbf{P}_2(\mathbf{Q})$, where $x, y, z \in \mathbf{R}$ for $\mathbf{P}_2(\mathbf{R})$, and where $x, y, z \in \mathbf{C}$ for $P_2(\mathbf{C})$.

(1.3) Remarks. A line C_F in $\mathbf{P}_2$ is the locus of all $w{:}x{:}y$ satisfying the equation $F(w, x, y) = aw + bx + cy = 0$. The line at infinity L_∞ is given by the equation $w = 0$. A point in $\mathbf{P}_2 - L_\infty$ has the form $1{:}x{:}y$ after multiplying with the factor w^{-1}. The point $1{:}x{:}y$ in the projective plane corresponds to (x, y) in the usual Cartesian plane. For a line L given by $aw + bx + cy = 0$ and L' given by $a'w + b'x + c'y = 0$ we have $L = L'$ if and only if $a{:}b{:}c = a'{:}b'{:}c'$ in the projective plane. In particular the points $a:b:c$ in the projective plane can be used to parametrize the lines in the projective plane.

From the theory of elimination of variables in beginning algebra we have the following geometric assertions of projective geometry whose verification is left to the reader.

(1.4) Assertion. Two distinct points P and P' in $\mathbf{P}_2(\mathbf{C})$ lie on a unique line L in the projective plane, and, further, if P and P' are rational points, then the line L is rational. Two distinct lines L and L' in $\mathbf{P}_2(\mathbf{C})$ intersect at a unique point P, and, further, if L and L' are rational lines, then the intersection point P is rational.

The projective line L with equation $L: aw + by + cy = 0$ determines the line $a + bx + cy = 0$ in the Cartesian plane. Two projective lines $L: aw + bx + cy = 0$ and $L': a'w + b'x + c'y = 0$ intersect on the line at infinity $w = 0$ if and only if $b:c = b':c'$, that is, the pairs (b, c) and (b', c') are proportional. The corresponding lines in the x,y-plane given by

$$a + bx + cy = 0 \quad \text{and} \quad a' + b'x + c'y = 0$$

have the same slope or are parallel exactly when the projective lines intersect at infinity. Now the reader is invited to reconsider the correspondence between rational points on two rational lines L and L' which arises by intersecting L and L' with all rational lines through a fixed point P_0 not on either L or L'.

To define plane curves in projective space, we use nonzero homogeneous polynomials $F(w, x, y) \in \mathbf{C}[w, x, y]$. Then we have the relation $F(qw, qx, qy) = q^d F(w, x, y)$, where $q \in \mathbf{C}$ and d is the degree of the homogenous polynomial $F(w, x, y)$. The locus C_F is the set of all $w:x:y$ in the projective plane such that $F(w, x, y) = 0$. Again the complex points of C_F are denoted by $C_F(\mathbf{C}) \subset \mathbf{P}_2(\mathbf{C})$, and, moreover, $C_F(\mathbf{C}) = C_{F'}(\mathbf{C})$ if and only if $F(w, x, y)$ and $F'(w, x, y)$ are proportional with a nonzero complex number. This assertion is not completely evident and is taken up again in Chapter 2.

(1.5) Definition. A rational (resp. real) plane curve in $\mathbf{P}_2$ is one of the form C_F where $F(w, x, y)$ has rational (resp. real) coefficients.

As in the x,y-plane for a rational plane curve C_F, we have rational, real, and complex points $C_F(\mathbf{Q}) \subset C_F(\mathbf{R}) \subset C_F(\mathbf{C})$.

(1.6) Remark. The above definition of a rational plane curve is an arithmetic notion, and it means the curve is defined over $\mathbf{Q}$. There is a geometric concept of rational curve (genus = 0) which should not be confused with (1.5). Geometric rationality is defined in terms of the equation of the curve $f(x, y) = 0$ over k and the field of fractions of $k[x, y]/(f)$.