Anthony W. Knapp

Elliptic Curves

椭圆曲线

PRINCETON UNIVERSITY PRESS

世界图书出版公司
www.wpcbj.com.cn

Elliptic Curves

by

Anthony W. Knapp

Mathematical Notes 40

PRINCETON UNIVERSITY PRESS

PRINCETON, NEW JERSEY
1992

图书在版编目（CIP）数据

椭圆曲线 = Elliptic Curves：英文/（美）奈普（Knapp，A. W.）著 . —影印本 . —北京：世界图书出版公司北京公司，2012. 8
ISBN 978 – 7 – 5100 – 5066 – 4

Ⅰ. ①椭…　Ⅱ. ①奈…　Ⅲ. ①椭圆曲线—英文　Ⅳ. ①O187. 1

中国版本图书馆 CIP 数据核字（2012）第 186120 号

书　　名：	Elliptic Curves
作　　者：	Anthony W. Knapp
中 译 名：	椭圆曲线
责任编辑：	高蓉　刘慧
出 版 者：	世界图书出版公司北京公司
印 刷 者：	三河市国英印务有限公司
发　　行：	世界图书出版公司北京公司（北京朝内大街 137 号 100010）
联系电话：	010 – 64021602，010 – 64015659
电子信箱：	kjb@ wpcbj. com. cn
开　　本：	24 开
印　　张：	18. 5
版　　次：	2013 年 1 月
版权登记：	图字：01 – 2012 – 5579
书　　号：	978 – 7 – 5100 – 5066 – 4　　**定　　价：** 69. 00 元

Elliptic Curves

Princeton University Press books are printed on acid-free paper, and meet the guidelines for permanence and durability of the Committee on Production Guidelines for Book Longevity of the Council on Library Resources

Library of Congress Cataloging-in-Publication Data

Knapp, Anthony W.
Elliptic curves / by Anthony W. Knapp.
p. cm.—(Mathematical notes ; 40)
Includes bibliographical references and index.
ISBN 0-691-08559-5 (PB)
1. Curves, Elliptic. I. Title. II. Series: Mathematical notes (Princeton University Press) ; 40.
QA567.2.E44K53 1993 516.3'52—dc20 92-22183

To Susan

PREFACE

For introductory purposes, an elliptic curve over the rationals is an equation $y^2 = P(x)$, where P is a monic polynomial of degree three with rational coefficients and with distinct complex roots.

The points on such a curve, together with a point at infinity, form an abelian group under a geometric definition of addition. Namely if we take two points on the curve and connect them by a line, the line will intersect the curve in a third point. The reflection of that third point in the x-axis is taken as the sum of the given points. The identity is the point at infinity. According to Mordell's Theorem, the abelian group of points on the curve with rational coordinates is finitely generated. A theorem of Lutz and Nagell describes the torsion subgroup completely, but the rank of the free abelian part is as yet not fully understood.

This is the essence of the basic theory of rational elliptic curves. The first five of the twelve chapters of this book give an account of this theory, together with many examples and number-theoretic applications. This is beautiful mathematics, of interest to people in many fields. Except for one small part of the proof of Mordell's Theorem, it is elementary, requiring only undergraduate mathematics. Accordingly the presentation avoids most of the machinery of algebraic geometry.

A related theory concerns elliptic curves over the complex numbers, or Riemann surfaces of genus one. This subject requires complex variable theory and is discussed in Chapter VI. It leads naturally to the topic of modular forms, which is the subject of Chapters VIII and IX.

But the book is really about something deeper, the twentieth-century discovery of a remarkable connection between automorphy and arithmetic algebraic geometry. This connection first shows up in the coincidence of L functions that arise from some very special modular forms ("automorphic" L functions) with L functions that arise from number theory ("arithmetic" or "geometric" L functions, also called "motivic"). Chapter VII introduces this theme. The automorphic L functions have manageable analytic properties, while the arithmetic L functions encode subtle number-theoretic information. The fact that the arithmetic L functions are automorphic enables one to bring a great deal of mathematics to bear on extracting the number-theoretic information from the L function.

The prototype for this phenomenon is the Riemann zeta function $\zeta(s)$, which should be considered as an arithmetic L function defined initially for Re $s > 1$. An example of subtle number-theoretic information that $\zeta(s)$ encodes is the Prime Number Theorem, which follows from the nanvanishing of $\zeta(s)$ for Re $s = 1$. In particular, this property of $\zeta(s)$ is a property of points s outside the initial domain of $\zeta(s)$. To get a handle on analytic properties of $\zeta(s)$, one proves that $\zeta(s)$ has an analytic continuation and a functional equation. These properties are completely formal once one establishes a relationship between $\zeta(s)$ and a theta function with known transformation properties. Establishing this relationship is the same as proving that $\zeta(s)$ is an automorphic L function.

The main examples of Chapter VII are the Dirichlet L functions $L(s,\chi)$. These too are arithmetic L functions defined initially for Re $s > 1$. They encode Dirichlet's Theorem on primes in arithmetic progressions, which follows from the nonvanishing of all $L(s,\chi)$ at $s = 1$. As with $\zeta(s)$, the relevant properties of $L(s,\chi)$ are outside the initial domain. Also as with $\zeta(s)$, one gets at the analytic continuation and functional equation of $L(s,\chi)$ by identifying $L(s,\chi)$ with an automorphic L function.

The examples at the level of Chapter VII are fairly easy. Further examples, generalizing the Dirichlet L functions in a natural way, arise in abelian class field theory, are well understood even if not easy, and will not be discussed in this book. The simplest L functions that are not well understood come from elliptic curves. An elliptic curve has a geometric L function $L(s,E)$ initially defined for Re $s > \frac{3}{2}$. An example conjecturally of the subtle information that $L(s,E)$ encodes is the rank of the free abelian group of rational points on the curve. This rank is believed to be the order of vanishing of $L(s,E)$ at $s = 1$. Once again, the relevant property of $L(s,E)$ is outside the initial domain. To address the necessary analytic continuation, one would like to know that $L(s,E)$ is an automorphic L function. Work of Eichler and Shimura provides a clue where to look for such a relationship. Eichler and Shimura gave a construction for passing from certain cusp forms of weight two for Hecke subgroups of the modular group to rational elliptic curves. Under this construction, the L function of the cusp form (which is an automorphic L function) equals the L function of the elliptic curve. The Taniyama-Weil Conjecture expects conversely that every elliptic curve arises from this construction, followed by a relatively simple map between elliptic curves. This conjecture appears to be very deep; a theorem of Frey, Serre, and Ribet says that it implies Fermat's Last Theorem. The final three chapters discuss these matters; the last two take for granted more mathematics than do the earlier chapters.

If the theme were continued beyond the twelve chapters that are here, eventually it would lead to the Langlands program, which brings in representation theory on the automorphic side of this correspondence. As a representation theorist, I come to elliptic curves from the point of view of the Langlands program. Although the book neither uses nor develops any representation theory, elliptic curves do give the simplest case of the program where the correspondence of L functions is not completely understood. Furthermore representation-theoretic methods occasionally yield results about elliptic curves that seem inaccessible by classical methods. From my point of view, they are an appropriate place to begin to study and appreciate the Langlands program. A beginning guide to the literature in this area appears in the section of Notes at the end of the book.

This book grew out of a brilliant series of a half dozen lectures by Don Zagier at the Tata Institute of Fundamental Research in Bombay in January 1988. The book incorporates notes from parts of courses that I gave at SUNY Stony Brook in Spring 1989 and Spring 1990. The organization owes a great deal to Zagier's lectures, and I have reproduced a number of illuminating examples of his. I am indebted to Zagier for offering his series of lectures.

Much of the mathematics here can already be found in other books, even if it has not been assembled in quite this way. Some sections of this book follow sections of other books rather closely. Notable among these other books are Fulton [1989],* Hartshorne [1977], Husemoller [1987], Lang [1976] and [1987], Ogg [1969a], Serre [1973a], Shimura [1971a], Silverman [1986], and Walker [1950]. The expository paper Swinnerton-Dyer and Birch [1975] was also especially helpful. Detailed acknowledgments of these dependences may be found in the section of Notes at the end.

In addition, I would like to thank the following people for help in various ways, some large and some small: H. Farkas, N. Katz, S. Kudla, R. P. Langlands, S. Lichtenbaum, H. Matumoto, C.-H. Sah, V. Schechtman, and R. Stingley. The type-setting was by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX, and the Figures were drawn with Mathematica®. Financial support in part was from the National Science Foundation in the form of grants DMS 87-23046 and DMS 91-00367.

A. W. Knapp
January, 1992

*A name followed by a bracketed year is an allusion to the list of References at the end of the book. The date is followed by a letter in case of ambiguity.

STANDARD NOTATION

Item	Meaning
$\#S$ or $\lvert S\rvert$	number of elements in S
$\emptyset$	empty set
A^c	A complement
n positive	$n > 0$
$\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$	integers, rationals, reals, complex numbers
$\operatorname{Re} z$	real part of z
$\operatorname{Im} z$	imaginary part of z
$O(1)$	bounded term
$o(1)$	term tending to 0
$\doteq$	approximately numerically equal
$\sim$	asymptotic to, with ratio tending to 1
$\mathbf{Z}_m$	integers modulo m
$a \equiv b \bmod m$	m divides $a - b$
$a \mid b$	a divides b
$\mathrm{GCD}(a, b)$	greatest common divisor
1	multiplicative identity
1 or I	identity matrix
$\dim V$	dimension of vector space
V'	dual of vector space
$\mathrm{End}_k(V)$	linear maps of vector space to itself
$GL(n, k)$	general linear group over a field k
$SL(2, \mathbf{R})$, $SL(2, \mathbf{Z})$	group of 2-by-2 matrices of determinant 1
$\operatorname{Tr} A$	trace of A
A^{tr}	transpose of A
$R^\times$	multiplicative group of invertible elements
$\bar{k}$	algebraic closure of k
$[A : B]$	index of B in A, or degree of A in B
$\mathrm{Aut}_k(K)$	automorphism group of K fixing k
$\sum \oplus$	direct sum (for emphasis)
π_1	fundamental group

CONTENTS

Elliptic Curves

CHAPTER I

OVERVIEW

Diophantus lived in Alexandria around 250 A.D. He published a series of books, *Arithmetica*, in 13 volumes, which were lost for more than a thousand years. They were found again about 1570, and in the next century Fermat studied a translation. Despite the passage of so much time, much of Diophantus's work had still not been rediscovered. The books are in the style of problems and solutions. The early volumes introduced, apparently for the first time, algebraic notation and equations, as well as negative numbers. Later volumes dealt with number theory. The work of Diophantus is so stunning that equations to be solved in number theory are often called Diophantine equations in his honor.

Basic Problem (affine). We consider the locus $f(x, y) = 0$ with f a nonzero polynomial. To fix the ideas, think of $\mathbb{Q}$ coefficients and of solutions with x and y in $\mathbb{Q}$. Clearing denominators, we can always adjust f so that we have $\mathbb{Z}$ coefficients and are seeking $\mathbb{Q}$ solutions, i.e., solutions with x and y in $\mathbb{Q}$.

Basic Problem (projective). We study the locus $F(x, y, w) = 0$ with F a nonzero homogeneous polynomial of some degree d, and we seek "projective" $\mathbb{Q}$ solutions. This means we identify solutions (x, y, w) and $(\lambda x, \lambda y, \lambda w)$ for $\lambda \neq 0$ and we discard $(0, 0, 0)$. Rational solutions automatically give us integer solutions, by clearing denominators.

The two problems are related, in that we can pass from each to the other. Two examples will illustrate.

EXAMPLE 1. Fermat equation $x^d + y^d = z^d$. This equation is projective. We can reduce to the affine case by dividing by z^d: $u^d + v^d = 1$. Relatively prime integer solutions of $x^d + y^d = z^d$ (with $z \neq 0$) correspond to rational solutions of $u^d + v^d = 1$ provided that we identify a solution (x, y, z) with its negative $(-x, -y, -z)$. This process of reducing to the affine case loses "affine solutions at ∞," those with $z = 0$, but they are trivial here. **Fermat's Last Theorem** is the conjecture that the equation has no nontrivial solutions for $d > 2$.

EXAMPLE 2. The affine equation $y^2 = x^3 + 1$. This becomes a projective equation if we put $y = \frac{v}{w}$ and $x = \frac{u}{w}$. We get $wv^2 = u^3 + w^3$. (Or we simply insert powers of w to make all terms cubic: $wy^2 = x^3 + w^3$.)

Diophantus considered the affine problem in degrees 1, 2, and 3.

Case of degree 1.

An affine line is of the form

$$ax + by + cz = 0 \qquad \text{with } a \text{ and } b \text{ not both } 0.$$

Two such lines are the same if and only if the coefficients of one are a multiple of the coefficients of the other.

A projective line is of the form

$$ax + by + cw = 0 \qquad \text{with } a, b, c \text{ not all } 0.$$

Again, two such lines are the same if and only if the coefficients of one are a multiple of the coefficients of the other. The line with $a = b = 0$ and $c = 1$ is $w = 0$, which is the "line at ∞."

Two facts are clear:

1) $\mathbb{Q}$ solutions exist (projectively).
2) We can parametrize all $\mathbb{Q}$ solutions.

Case of degree 2.

First we consider this case projectively. Then it turns out to be possible to make a linear change of variables so that the equation is

$$ax^2 + by^2 + cw^2 = 0. \tag{1.1}$$

The questions we shall address are:

(I) Do (nonzero) solutions exist?
(II) If solutions exist, how are they parametrized?

Of these, only the second question was considered by Diophantus.

We shall return to the second question shortly. We begin with a discussion of Question I, first giving two examples.

EXAMPLE 1. $x^2 + y^2 + z^2 = 0$ has no $\mathbb{Q}$ solutions since it has no $\mathbb{R}$ solutions.

EXAMPLE 2. $x^2 + y^2 = 3z^2$ has no $\mathbb{Q}$ solutions since it has no solutions modulo 9 in which some variable is not divisible by 3. This condition is necessary since existence of a $\mathbb{Q}$ solution implies existence of a relatively prime $\mathbb{Z}$ solution. To see that the condition fails, we argue as follows: Modulo 3, the equation is $x^2 + y^2 \equiv 0$, which forces $x \equiv y \equiv 0$ mod 3. Then 3^2 divides $x^2 + y^2$ and so 3 divides z. Thus 3 divides all of x, y, and z, a contradiction.