LONDON MATHEMATICAL SOCIETY STUDENT TEXTS

Managing editor: Professor C.M. Series, Mathematics Institute
University of Warwick, Coventry CV4 7AL, United Kingdom

3 Local fields, J.W.S. CASSELS
4 An introduction to twistor theory: Second edition, S.A. HUGGETT & K.P. TOD
5 Introduction to general relativity, L.P. HUGHSTON & K.P. TOD
7 The theory of evolution and dynamical systems, J. HOFBAUER & K. SIGMUND
8 Summing and nuclear norms in Banach space theory, G.J.O. JAMESON
9 Automorphisms of surfaces after Nielsen and Thurston, A. CASSON & S. BLEILER
11 Spacetime and singularities, G. NABER
12 Undergraduate algebraic geometry, MILES REID
13 An introduction to Hankel operators, J.R. PARTINGTON
15 Presentations of groups: Second edition, D.L. JOHNSON
17 Aspects of quantum field theory in curved spacetime, S.A. FULLING
18 Braids and coverings: selected topics, VAGN LUNDSGAARD HANSEN
19 Steps in commutative algebra, R.Y. SHARP
20 Communication theory, C.M. GOLDIE & R.G.E. PINCH
21 Representations of finite groups of Lie type, FRANÇOIS DIGNE & JEAN MICHEL
22 Designs, graphs, codes, and their links, P.J. CAMERON & J.H. VAN LINT
23 Complex algebraic curves, FRANCES KIRWAN
24 Lectures on elliptic curves, J.W.S. CASSELS
25 Hyperbolic geometry, BIRGER IVERSEN
26 An introduction to the theory of L-functions and Eisenstein series, H. HIDA
27 Hilbert Space: compact operators and the trace theorem, J.R. RETHERFORD
28 Potential theory in the complex plane, T. RANSFORD
29 Undergraduate commutative algebra, M. REID
31 The Laplacian on a Riemannian manifold, S. ROSENBERG
32 Lectures on Lie groups and Lie algebras, R. CARTER, G. SEGAL & I. MACDONALD
33 A primer of algebraic *D*-modules, S.C. COUTINHO
34 Complex algebraic surfaces, A. BEAUVILLE
35 Young tableaux, W. FULTON
37 A mathematical introduction to wavelets, P. WOJTASZCZYK
38 Harmonic maps, loop groups and integrable systems, M. GUEST
39 Set theory for the working mathematician, K. CIESIELSKI
40 Ergodic theory and dynamical systems, M. POLLICOTT & M. YURI
41 The algorithmic resolution of diophantine equations, N.P. SMART
42 Equilibrium states in ergodic theory, G. KELLER

London Mathematical Society Student Texts 41

The Algorithmic Resolution of Diophantine Equations

Nigel P. Smart
Hewlett-Packard Laboratories, Bristol

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521641562

First published 1998

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

Smart, N. P. (Nigel Paul), 1967–
The algorithmic resolution of diophantine equations / N.P. smart
p. cm.
Includes bibliographical references and index.
ISBN 052164156X. – ISBN 0521646332 (pbk.)
1. Diophantine equations. I. Title.
QA242.S69 1998
512'. 72–dc21 98–24736 CIP

ISBN 978-0-521-64156-2 hardback
ISBN 978-0-521-64633-8 paperback

Transferred to digital printing 2007

To Maggie, Eleanor and Oliver.

Contents

Preface xi
Outline xii
Computer packages xv
Notation xv
Thanks xvi

Chapter I. Introduction 1
I.1. A brief history 2
I.2. Algorithms 7
I.3. What is a diophantine equation? 9
I.4. An elliptic curve 10

Part 1. Basic solution techniques 15

Chapter II. Local methods 17
II.1. p-adic numbers 17
II.2. p-adic numerical analysis 23
II.3. Exercises 32

Chapter III. Applications of local methods to diophantine equations 33
III.1. Applications of Strassmann's theorem 33
III.2. Skolem's method 36
III.3. The Hasse principle 39
III.4. Finding small solutions 40
III.5. Exercises 43

Chapter IV. Ternary quadratic forms 45
IV.1. A normal form 45
IV.2. Local solubility 46
IV.3. Global solubility 49
IV.4. New solutions for old 53
IV.5. Exercises 56

Chapter V. Computational diophantine approximation 59
V.1. Continued fractions 59
V.2. Approximation lattices 64
V.3. Lattices 65
V.4. The LLL–algorithm 71

V.5. Exercises 75

Chapter VI. Applications of the LLL–algorithm 77
VI.1. A 'fun' application 77
VI.2. Knapsack problems 79
VI.3. Approximating linear forms 82
VI.4. p-adic analogues 87
VI.5. Exercises 93

Part 2. Methods using linear forms in logarithms 95

Chapter VII. Thue equations 97
VII.1. Thue equations 97
VII.2. $X^4 - 2Y^4 = \pm 1$ 105
VII.3. The method of Bilu and Hanrot 108
VII.4. Integral points on elliptic curves (I) 111
VII.5. $Y^2 = X^3 - 6X - 14$ 113
VII.6. Exercises 116

Chapter VIII. Thue–Mahler equations 117
VIII.1. Thue–Mahler equations 117
VIII.2. The prime ideal removing lemma 118
VIII.3. The method 119
VIII.4. $X^3 - X^2Y + XY^2 + Y^3 = \pm 11^s$ 124
VIII.5. Exercises 132

Chapter IX. S-unit equations 133
IX.1. S-unit equations 133
IX.2. Sieving 141
IX.3. An S-unit equation in a cyclic quintic field 146
IX.4. Integral points on elliptic curves (II) 150
IX.5. Other applications 151
IX.6. Exercise 152

Chapter X. Triangularly connected decomposable form equations 153
X.1. Triangularly connected linear forms 153
X.2. TCDF equations 155
X.3. Solving TCDF equations 156
X.4. Exercises 163

Chapter XI. Discriminant form equations 165
XI.1. Discriminant and index forms 165
XI.2. The general case: discriminant forms as TCDFs 167
XI.3. A discriminant form equation in a cyclic quintic field 169
XI.4. Special cases 170
XI.5. Exercises 174

Part 3. Integral and rational points on curves 175

Chapter XII. Rational points on elliptic curves 177
XII.1. Basics on elliptic curves 177
XII.2. The weak Mordell–Weil theorem 182
XII.3. The Mordell–Weil theorem 190
XII.4. A conditional algorithm 192
XII.5. Exercises 194

Chapter XIII. Integral points on elliptic curves 197
XIII.1. Elliptic logarithms 197
XIII.2. Elliptic integrals and the AGM 198
XIII.3. Integral points 202
XIII.4. Integral points on the curve $Y^2 = X^3 - 2$ 206
XIII.5. S–integral points 207
XIII.6. Other methods and problems 210
XIII.7. Exercises 211

Chapter XIV. Curves of genus greater than one 213
XIV.1. Curves and their Jacobians 213
XIV.2. Hyperelliptic curves and their Jacobians 215
XIV.3. Rational points on curves of genus greater than one 217
XIV.4. Integral points on hyperelliptic and superelliptic curves 219
XIV.5. Fermat curves 221
XIV.6. Catalan's equation 222
XIV.7. Exercises 224

Appendix A. Linear forms in logarithms 225
A.1. Linear forms in complex logarithms 225
A.2. Linear forms in p-adic logarithms 225
A.3. Linear forms in elliptic logarithms 226

Appendix B. Two useful lemmata 229

References 231

Index 241

Preface

Many books have been devoted to the theoretical study of diophantine equations, an observation which should come as no surprise given that the study of such equations dates back over two thousand years. In theoretical work one is interested in determining the structure of the solution set to some equation. Is the set finite or infinite? Can one give an effective procedure to determine all the solutions? Do the solutions form a group of some sort? How are the rational solutions distributed amongst the real solutions? The list of questions that one can ask is endless.

In this book we shall concentrate on algorithms and methods for writing down all the solutions to an equation (if there are finitely many) or for determining explicitly the structure of all of the solutions (if there are infinitely many). Despite the long and noble career of diophantine equations, there appear to be only two books solely devoted to the study of explicit methods for their solution, namely Mordell's *Diophantine Equations* [**138**] and de Weger's *Algorithms For Diophantine Equations* [**208**].

Mordell's book gives a variety of techniques for solving various diophantine equations. However, sometimes he deals just with special cases and sometimes with general cases. Mordell does not concentrate on algorithmic questions and hence some of his methods appear at first sight to be recipes which only apply to certain special cases. This is not surprising as it was originally published in 1969, before the advent of the modern desktop computer.

The second book is de Weger's thesis, in which the systematic use of the LLL–algorithm was proposed for solving diophantine equations. This has revolutionized the subject and led to a great explosion in the number of papers devoted to algorithms for diophantine equations. De Weger's book was published in 1989 at the beginning of this revolution and it therefore only barely touches, for instance, on the algorithm for Thue–Mahler equations developed by Tzanakis and de Weger.

There have been many books which have studied diophantine equations from a theoretical standpoint, most notably the book by Shorey and Tijdeman [**167**], which gives an excellent account of the applications of Baker's theory of linear forms in logarithms.

The advent of modern computer technology has led to a number of books on algorithms for number theory. We shall require the use of various algorithms to solve problems in algebraic number theory. In particular we shall

require the solution of various problems in algebraic number fields, such as unit and class group computation. Many of the number field algorithms we require can be found in the books by Cohen [**32**], Pohst [**154**] and Pohst and Zassenhaus [**155**]. For up to date information one should perhaps consult the various conference proceedings, such as [**1**], [**33**], [**137**] and [**152**].

Therefore the time seems ripe for a new book on the computational side of this area. We shall aim to provide a coherent account of some of the many methods that can be used to find all the solutions to certain diophantine equations. However, we shall mainly be interested in methods which apply to a wide class of equations rather than just a few special examples. In some sense this is still a recipe book, but we hope a recipe book which gives the chef a range of skills for coping with a number of dishes.

We shall assume that you are familiar with standard undergraduate algebraic number theory up to, say, Dirichlet's units theorem and the finiteness of the class group. The book by Rose [**160**] covers most of what we will require bar the two aforementioned results on the units and class number. For these last two results you should perhaps consult another textbook such as that by Stewart and Tall [**187**].

We shall also assume that you have begun to study the more advanced theory that one meets as a graduate student, such as local fields. We shall, however, give a brief overview of the theory of local fields at the start. We shall furthermore take it that the reason you are reading this book is that you are interested in computations in number theory. This is not, therefore, a theoretical book but a practical one.

Outline

The book is divided into three parts. Part 1 will involve the study of the basic techniques which are used over and over again in solving diophantine equations. These are chiefly: the theory of p-adic numbers; the use of curves of genus zero; and the application of the algorithm of Lenstra, Lenstra and Lovász.

We shall, in Chapter II, start by giving a brief overview of the theory of p-adic numbers and local fields. Those of you who have not met local fields before should consult one of the excellent textbooks in this area. This subject is covered very well in other books such as the one by Cassels, [**24**], so we shall just review the main results we shall need. An explanation of p-adic analogues of results from numerical analysis will start our investigation.

In Chapter III we shall focus on how we can use the theory of p-adic numbers to solve diophantine equations, with the use of Hensel's lemma and Skolem's method. Finally we shall end our discussion of local fields with a brief discussion on how one can put all the local information together in various ways using Hasse's principle and sieving. The discussion of these

applications of local methods will be brief, as they are covered elsewhere, for instance in the book by Mordell [**138**] mentioned previously.

We shall then turn, in Chapter IV, to the discussion of the solution of ternary quadratic forms. These can also be characterized as curves in $\mathbb{P}^2$ of genus zero. Such equations are not only important in their own right but also occur in algorithms for solving more complicated equations, the idea being that ternary quadratic forms are 'easy' and if we can reduce our study of a hard equation to a set of ternary quadratic forms then we would have made life easier. The reason ternary quadratic forms are considered easy is that they all satisfy the Hasse principle.

In Chapter V we study the LLL–algorithm of Lenstra, Lenstra and Lovász. We shall develop the algorithm from scratch and go on to show how one can use it to give a lower bound on the size of the smallest non-zero vector in a lattice and the smallest distance between a given non-lattice vector and a vector in the lattice.

Chapter VI will be concerned with the application of the LLL–algorithm to various problems. For instance we shall see how one can use LLL to solve certain types of knapsack problem. We shall end by showing how one can use LLL to study problems of linear forms in real, complex and p-adic numbers. This technique is the one we shall use to reduce the stratospherically large bounds which arise from the theory of linear forms in logarithms.

Part 2 will be devoted to problems to which one can apply the theory of linear forms in logarithms of algebraic numbers and its generalizations to p-adic logarithms.

Thue and Thue–Mahler equations are dealt with in Chapters VII and VIII. Thue equations are equations of the form

$$F(X,Y) = m$$

whilst Thue–Mahler equations are of the form

$$F(X,Y) = mp_1^{z_1} \cdots p_t^{z_t},$$

where, in both cases, $F(X,Y)$ is a binary form of degree greater than three and m is some fixed integer. But in the case of Thue–Mahler equations we have the added complication of some given prime numbers p_i with some unknown exponents z_i. Thue equations have a finite number of integer solutions, whilst Thue–Mahler equations have infinitely many solutions which can be divided into finitely many families. Thue and Thue–Mahler equations form the easiest examples of classes of equations which can be dealt with by Baker's methods followed by a reduction process based on the LLL–algorithm. Chapter VII will conclude with an example of how to use Thue equations to solve another diophantine problem: that of finding all the integer points on an elliptic curve. This will be the first of three such methods we give to solve this problem.

The trick in solving Thue and Thue–Mahler equations is to reduce the problem to the study of S-unit equations, that is, equations of the form

$$\alpha_1\tau_1 + \alpha_2\tau_2 + 1 = 0,$$

where α_i are two fixed algebraic numbers and τ_1 and τ_2 are allowed to range over two finitely generated multiplicative subgroups of the algebraic numbers. Such an equation has only finitely many solutions, which we shall give an effective proof of in Chapter IX. Indeed we shall give an algorithm which can often be used in practice to solve such an equation. We end this chapter by showing how one could use an algorithm to solve S-unit equations to give another method for finding all the integral points on an elliptic curve.

In that late 1970s and early 1980s Győry showed how a very large set of diophantine equations could be reduced to the study of S-unit equations. This set of triangularly connected decomposable form equations (TCDF equations for short) is studied in Chapter X. These equations are a natural generalization of the Thue and Thue–Mahler equations considered earlier.

In Chapter XI we shall pay particular attention to a special type of TCDF equation, the set of discriminant form equations. We shall end this chapter by showing how discriminant form equations related to quartic number fields can be solved by using a combination of Thue equations and ternary quadratic forms, which bypasses the need to consider them as TCDF equations.

In Part 3 we shall consider methods for finding integral and rational solutions to curves such as elliptic, hyperelliptic and superelliptic equations.

In Chapters XII and XIII we shall concentrate on elliptic curves. It has been known for over 100 years that the set of rational points on an elliptic curve forms a group. Chapter XII will be devoted to giving an (almost) algorithmic proof of the result of Mordell that such a group is finitely generated. That there is no such algorithmic answer in general is due to the failure of the Hasse principle for curves of genus one. In Chapter XIII we shall use the method for determining generators of the group of rational points to give a third method for finding all the integral points on such a curve.

In Chapter XIV we shall look at recent work on generalizations of the methods for elliptic curves to curves of higher genus. In particular we shall concentrate on hyperelliptic curves. Owing to Faltings proof of the Mordell conjecture we now know that there are only finitely many rational points on a curve of genus greater than one. However, at present there are only ad hoc techniques to find all the rational points in any given example. We shall present a quick overview of some of the work done in this area and its link with the Jacobian variety of a curve of genus greater than one.

In this final chapter we shall also cover a few odds and ends which we have not covered in other chapters. In particular no book on diophantine equations would be complete without a passing mention of Wiles' proof of Fermat's last theorem. In this last section we shall describe the link between Fermat's last theorem and elliptic curves, although we shall not go into any

details, as that would involve going into the theory of modular functions and Galois representations. In addition in this last chapter we shall look at the ABC conjecture, which is in some sense a generalization of the two–term S-unit equations which are met elsewhere in the book.

Clearly we have not even attempted to cover all the different types of equation that can be studied. Nor have we covered much of the extensive theoretical work on diophantine equations. The subjects chosen are a personal choice, as is fitting for a recipe book. There are some topics which we have left out owing to lack of space.

One is the application of Padé approximations and hypergeometric functions. In this work instead of trying to approximate linear forms in logarithms one looks at, for example, approximating numbers of the form

$$\sqrt[n]{1+a}.$$

Readers interested in following up such work should consult [**159**], [**29**], [**120**], [**206**], [**8**] and [**9**].

Another is the algorithmic study of diophantine properties of linear recurrence sequences. If this area is what interests you then why not start by looking at [**145**] and [**214**]. A good introduction to this area can be found in the relevant chapters of [**167**].

Computer packages

There are currently many computer packages for performing number theoretic calculations. We could be content with just using one of the main computer algebra packages such as **Maple** or **Mathematica**. However, we shall need to be able to compute units and class groups of number fields, etc. Hence access to a package like **PARI** [**7**], **KANT** [**42**], **SIMATH** [**177**], **LiDIA** [**122**] or **MAGMA** [**17**], would seem desirable.

Many of the examples in this book were carried out with the aid of a computer, so you should not expect to be able to follow an example through by hand (except in some easy cases). However, a computer can solve most of the examples in this book in a matter of seconds.

Notation

As usual we shall denote the complex, real and rational numbers by $\mathbb{C}, \mathbb{R}$ and $\mathbb{Q}$. The ring of integers we shall denote by $\mathbb{Z}$ while the set of non-negative integers will be denoted by $\mathbb{N}$. Multiplication of numbers will denoted by $6 = 2 \cdot 3$, while a decimal point will be given by $2.3 = \frac{23}{10}$.

Some of the notation used could be considered non-standard in that not all authors use the same notation. To make this clear we spell out the possible non-standard notation now.

The notation $\mathbb{Z}_p$ will be reserved for the p-adic integers, the p-adic numbers being denoted by $\mathbb{Q}_p$. The set of integers modulo m will then be denoted by $\mathbb{Z}/m\mathbb{Z}$. The finite field of q elements will be denoted by $\mathbb{F}_q$

For a real number x, the symbol $\lfloor x \rfloor$ will denote the floor function, i.e. it returns the largest integer less than x. The symbol $\lceil x \rceil$ will denote the ceiling function, i.e. the smallest integer greater than x. The nearest integer function will be denoted $[x]$, with any fixed convention for numbers of the form $(2m+1)/2$. The symbol $\{x\}$ will be used to denote $|x - [x]|$.

For a complex number z the real and imaginary parts will be denoted by $\Re(z)$ and $\Im(z)$.

The symbol ${}^nC_r$ will denote the binomial coefficient

$$\frac{n!}{r!(n-r)!}.$$

The greatest common divisor of two integers a and b will be denoted by (a, b).

If K is a number field then we let $\mathcal{O}_K$ denote its maximal order. The unit and class groups of K will be denoted by $\mathcal{O}_K^*$ and CL_K respectively. If $\alpha_1, \ldots, \alpha_t$ are elements of $\mathcal{O}_K$, for some number field, K then we let $(\alpha_1, \ldots, \alpha_t)$ denote the ideal generated over $\mathcal{O}_K$ by $\alpha_1, \ldots, \alpha_t$.

All other notation will be defined as and when required.

Thanks

The author would like to thank J. Cremona, S. Crouch, E. Mansfield, A. Pethő, R. Shipsey, S. Siksek, N. Stephens and B. M. M. de Weger, who read various parts of the manuscript at various stages. Any mistakes are still, however, my own fault. None of the mathematics in this book is new and an attempt has been made to provide references to major results. If you feel that due credit has not been given for certain results, then accept the authors apologies in advance.

Finally, thanks are due to J. Cremona for a TEX macro for typesetting the algorithms.