Number Theory

I: Tools and Diophantine Equations

II: Analytic and Modern Methods

by

Henri COHEN

Version of October 28, 2006

This manuscript was typed using $\mathcal{A}_\mathcal{M}\mathcal{S}$-*LaTeX and the Springer-Verlag* `clmono01` *macro package.*

Preface

This book deals with several aspects of what is now called "explicit number theory," not including the essential algorithmic aspects, which are for the most part covered by two other books of the author [Coh0] and [Coh1]. The central (although not unique) theme is the solution of Diophantine equations, i.e., equations or systems of polynomial equations that must be solved in integers, rational numbers, or more generally in algebraic numbers. This theme is in particular the central motivation for the modern theory of arithmetic algebraic geometry. We will consider it through three of its most basic aspects.

The first is the *local* aspect: the invention of p-adic numbers and their generalizations by K. Hensel was a major breakthrough, enabling in particular the simultaneous treatment of congruences modulo prime powers. But more importantly, one can do *analysis* in p-adic fields, and this goes much further than the simple definition of p-adic numbers. The local study of equations is usually not very difficult. We start by looking at solutions in *finite fields*, where important theorems such as the Weil bounds and Deligne's theorem on the Weil conjectures come into play. We then *lift* these solutions to local solutions using *Hensel lifting*.

The second aspect is the *global* aspect: the use of number fields, and in particular of class groups and unit groups. Although local considerations can give a considerable amount of information on Diophantine problems, the "local to global" principles are unfortunately rather rare, and we will see many examples of failure. Concerning the global aspect, we will first require as a prerequisite of the reader that he or she is familiar with the standard basic theory of number fields, up to and including the finiteness of the class group and Dirichlet's structure theorem for the unit group. This can be found in many textbooks such as [Sam] and [Marc]. Second, and this is less standard, we will always assume that we have at our disposal a computer algebra system (CAS) that is able to compute rings of integers, class and unit groups, generators of principal ideals, and related questions. Such CAS are now very common, for instance `Kash`, `magma`, and `Pari/GP`, to cite the most useful in algebraic number theory.

The third aspect is the theory of zeta and L-functions. This can be considered as a *unifying theme*[1] for the whole subject, and embodies in a beautiful way the local and global aspects of Diophantine problems. Indeed, these functions are defined through the local aspects of the problems, but their analytic behavior is intimately linked to the global aspects. A first example is given by the Dedekind zeta function of a number field, which is defined only through the splitting behavior of the primes, but whose leading term at $s = 0$ contains at the same time explicit information on the unit rank, the class number, the

[1] expression due to Don Zagier

regulator, and the number of roots of unity of the number field. A second very important example, which is one of the most beautiful and important conjectures in the whole of number theory (and perhaps of the whole of mathematics), the Birch and Swinnerton-Dyer conjecture, says that the behavior at $s = 1$ of the L-function of an elliptic curve defined over $\mathbb{Q}$ contains at the same time explicit information on the rank of the group of rational points on the curve, on the regulator, and on the order of the torsion group of the group of rational points, in complete analogy with the case of the Dedekind zeta function. In addition to the purely *analytical* problems, the theory of L-functions contains beautiful results (and conjectures) on *special values*, of which Euler's formula $\sum_{n\geqslant 1} 1/n^2 = \pi^2/6$ is a special case.

This book can be considered as having four main parts. A first part gives the tools necessary for Diophantine problems: equations over finite fields, number fields, and finally local fields such as $\mathfrak{p}$-adic fields (Chapters 1, 2, 3, 4, and part of Chapter 5). The emphasis will be mainly on the theory of $\mathfrak{p}$-adic fields (Chapter 4), since the reader has probably less familiarity with these. Note that we will consider function fields only in Chapter 7, as a tool for proving Hasse's theorem on elliptic curves. An important tool that we will introduce at the end of Chapter 3 is the theory of the Stickelberger ideal over cyclotomic fields, together with the important applications to the Eisenstein reciprocity law, and the Davenport–Hasse relations. Through Eisenstein reciprocity this theory will enable us to prove Wieferich's criterion for the first case of Fermat's last theorem (FLT), and it will also be an essential tool in the proof of Catalan's conjecture given in Chapter 16.

A second part is the study of certain basic Diophantine equations or systems of equations (Chapters 5, 6, 7, and 8). It should be stressed that even though a number of general techniques are available, each Diophantine equation poses a new problem, and it is difficult to know in advance whether it will be easy to solve. Even without mentioning *families* of Diophantine equations such as FLT, the congruent number problem, or Catalan's equation, all of which will be stated below, proving for instance that a specific equation such as $x^3 + y^5 = z^7$ with x, y coprime integers has no solution with $xyz \neq 0$ seems presently out of reach, although it has been proved (based on a deep theorem of Faltings) that there are only finitely many solutions; see [Dar-Gra] and Chapter 14. Note also that it has been shown by Yu. Matiyasevich (after a considerable amount of work by other authors) in answer to Hilbert's tenth problem that there cannot exist a general algorithm for solving Diophantine equations.

The third part (Chapters 9, 10, and 11) deals with the detailed study of analytic objects linked to algebraic number theory: Bernoulli polynomials and numbers, the gamma function, and zeta and L-functions of Dirichlet characters, which are the simplest types of L-functions. In Chapter 11 we also study p-adic analogues of the gamma, zeta, and L-functions, which have come to play an important role in number theory, and in particular the Gross–

Koblitz formula for Morita's p-adic gamma function. In particular we will see that this formula leads to remarkably simple proofs of Stickelberger's congruence and the Hasse–Davenport product relation. More general L-functions such as Hecke L-functions for Grössencharacters, Artin L-functions for Galois representations, or L-functions attached to modular forms, elliptic curves, or higher-dimensional objects are mentioned in several places, but a systematic exposition of their properties would be beyond the scope of this book.

Much more sophisticated techniques have been brought to bear on the subject of Diophantine equations, and it is impossible to be exhaustive. Because the author is not an expert in most of these techniques they are not studied in the first three parts of the book. However considering their importance, I have asked a number of much more knowledgeable people to write a few chapters on these techniques, and I have written two myself, and this forms the fourth and last part of the book (Chapters 12 to 16). These chapters have a different flavor from the rest of the book: they are in general not self-contained, are of a higher mathematical sophistication than the rest, and usually have no exercises. Chapter 12, written by Yann Bugeaud, Guillaume Hanrot, and Maurice Mignotte, deals with the applications of Baker's explicit results on linear forms in logarithms of algebraic numbers, which permit the solution of a large class of Diophantine equations such as Thue equations and norm form equations, and includes some recent spectacular successes. Paradoxically, the similar problems on elliptic curves are considerably less technical, and are studied in detail in Section 8.7. Chapter 13, written by Sylvain Duquesne, deals with the search for rational points on curves of genus greater than or equal to 2, restricting for simplicity to the case of hyperelliptic curves of genus 2 (the case of genus 0, in other words of quadratic forms, is treated in Chapters 5 and 6, and the case of genus 1, essentially of elliptic curves, is treated in Chapters 7 and 8). Chapter 14, written by the author, deals with the so-called super-Fermat equation $x^p + y^q = z^r$, on which several methods have been used, including ordinary algebraic number theory, classical invariant theory, rational points on higher genus curves, and Ribet–Wiles type methods. The only proofs that are included are those coming from algebraic number theory. Chapter 15, written by Samir Siksek, deals with the use of Galois representations, and in particular of Ribet's level-lowering theorem and Wiles's and Taylor–Wiles's theorem proving the modularity conjecture. The main application is to equations of "abc" type, in other words, equations of the form $a+b+c = 0$ with a, b, and c highly composite, the "easiest" application of this method being the proof of FLT. The author of this chapter has tried to hide all the sophisticated mathematics and to present the method as a black box that can be used without completely understanding the underlying theory. Finally Chapter 16, also written by the author, gives the complete proof of Catalan's conjecture by P. Mihăilescu. It is entirely based on notes of Yu. Bilu, R. Schoof, and especially of J. Boéchat and M. Mischler, and the only reason that it is not self-contained is that it will be necessary to assume the validity of an important theorem of F. Thaine on the annihilator of the plus part of the class group of cyclotomic fields.

Warnings

Since several mathematical conventions and notation are not the same from one mathematical culture to the next, I have decided to use systematically unambiguous terminology, and when the notation clash, the French notation. Here are the most important:

– We will systematically say that a is strictly greater than b, or greater than or equal to b (or b is strictly less than a, or less than or equal to a), although the English terminology a is greater than b means in fact one of the two (I don't remember which one, and that is one of the main reasons I refuse to use it) and the French terminology means the other. Similarly, positive and negative are ambiguous (does it include the number 0)? Even though the expression "x is nonnegative" is slightly ambiguous, it is useful, and I *will* allow myself to use it, with the meaning $x \geqslant 0$.
– Although we will almost never deal with noncommutative fields (which is a contradiction in terms since in principle the word field implies commutativity), we will usually not use the word field alone. Either we will write explicitly commutative (or noncommutative) field, or we will deal with specific classes of fields, such as finite fields, $\mathfrak{p}$-adic fields, local fields, number fields, etc...., for which commutativity is clear. Note that the "proper" way in English-language texts to talk about noncommutative fields is to call them either skew fields or division algebras. In any case this will not be an issue since the only appearances of skew fields will be in Chapter 2, where we will prove that finite skew fields are commutative, and in Chapter 7 about endomorphism rings of elliptic curves over finite fields.
– The GCD (resp., the LCM) of two integers can be denoted by (a, b) (resp., by $[a, b]$), but to avoid ambiguities, I will systematically use the explicit notation $\gcd(a, b)$ (resp., $\operatorname{lcm}(a, b)$), and similarly when more than two integers are involved.
– An open interval with endpoints a and b is denoted by (a, b) in the English literature, and by $]a, b[$ in the French literature. I will use the French notation, and similarly for half-open intervals $(a, b]$ and $[a, b)$, which I will denote by $]a, b]$ and $[a, b[$. Although it is impossible to change such a well-entrenched notation, I urge my English-speaking readers to realize the dreadful ambiguity of the notation (a, b), which can either mean the ordered pair (a, b), the GCD of a and b, or the open interval.
– The trigonometric functions $\sec(x)$ and $\csc(x)$ do not exist in France, so I will not use them. The functions $\tan(x)$, $\operatorname{cotan}(x)$, $\cosh(x)$, $\sinh(x)$, $\tanh(x)$, and $\operatorname{cotanh}(x)$ are denoted respectively $\operatorname{tg}(x)$, $\operatorname{cotg}(x)$, $\operatorname{ch}(x)$,

sh(x), th(x), and coth(x) in France, but for once to bow to the majority I will use the English names.
– $\Re(s)$ and $\Im(s)$ denote the real and imaginary part of the complex number, the notation coming from the standard TEX macros.

Notation

In addition to the standard notation of number theory we will use the following notation.

– We will often use the practical self-explanatory notation $\mathbb{Z}_{>0}$, $\mathbb{Z}_{\geqslant 0}$, $\mathbb{Z}_{<0}$, $\mathbb{Z}_{\leqslant 0}$, and generalizations thereof, which avoid using long wordage. On the other hand I prefer not to use the notation $\mathbb{N}$ (for $\mathbb{Z}_{\geqslant 0}$, or is it $\mathbb{Z}_{>0}$?).
– If a and b are nonzero integers, we write $\gcd(a, b^\infty)$ for the limit of the ultimately constant sequence $\gcd(a, b^n)$ as $n \to \infty$. We have of course $\gcd(a, b^\infty) = \prod_{p \mid \gcd(a,b)} p^{v_p(a)}$, and $a/\gcd(a, b^\infty)$ is the largest divisor of a coprime to b.
– If n is a nonzero integer and $d \mid n$, we write $d \| n$ if $\gcd(d, n/d) = 1$. Note that this is *not* the same thing as the condition $d^2 \nmid n$, except if d is prime.
– If $x \in \mathbb{R}$, we denote by $\lfloor x \rfloor$ the largest integer less than or equal to x (the *floor* of x), by $\lceil x \rceil$ the smallest integer greater than or equal to x (the *ceiling* of x, which is equal to $\lfloor x \rfloor + 1$ if and only if $x \notin \mathbb{Z}$), and by $\lfloor x \rceil$ the nearest integer to x (or one of the two if $x \in 1/2 + \mathbb{Z}$), so that $\lfloor x \rceil = \lfloor x + 1/2 \rfloor$.
– For any α belonging to a field K of characteristic zero and any $k \in \mathbb{Z}_{\geqslant 0}$ we set

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!} .$$

In particular if $\alpha \in \mathbb{Z}_{\geqslant 0}$ we have $\binom{\alpha}{k} = 0$ if $k > \alpha$, and in this case we will set $\binom{\alpha}{k} = 0$ also when $k < 0$. On the other hand $\binom{\alpha}{k}$ is *undetermined* for $k < 0$ if $\alpha \notin \mathbb{Z}_{\geqslant 0}$.
– Capital italic letters such as K and L will usually denote number fields.
– Capital calligraphic letters such as $\mathcal{K}$ and $\mathcal{L}$ will denote general $\mathfrak{p}$-adic fields (for specific ones, we write for instance $K_{\mathfrak{p}}$).
– The capital letter $\mathbb{Z}$ indexed by a capital italic or calligraphic letter such as $\mathbb{Z}_K$, $\mathbb{Z}_L$, $\mathbb{Z}_{\mathcal{K}}$, etc.... will always denote the ring of integers of the corresponding field.
– Capital italic letters such as A, B, C, G, H, S, T, U, V, W, or lowercase italic letters such as f, g, h will usually denote polynomials or formal power series with coefficients in some base ring or field. The coefficient of degree m of these polynomials or power series will be denoted by the corresponding letter indexed by m, such as A_m, B_m, etc.... Thus we will always write (for

instance) $A(X) = A_d X^d + A_{d-1} X^{d-1} + \cdots + A_0$, so that the ith elementary symmetric function of the roots is equal to $(-1)^i A_{d-i}/A_d$.

Acknowledgments

A large part of the material on local fields has been taken with little change from the remarkable book by Cassels [Cas1], and also from unpublished notes of Jaulent written in 1994. For p-adic analysis, I have also liberally borrowed from work of Robert, in particular his superb GTM volume [Rob1]. For part of the material on elliptic curves I have borrowed from another excellent book by Cassels [Cas2], as well as the treatises of Cremona and Silverman [Cre2], [Sil1], [Sil2], and the introductory book by Silverman–Tate [Sil-Tat]. I have also borrowed from the classical books by Borevich–Shafarevich [Bor-Sha], Serre [Ser1], Ireland–Rosen [Ire-Ros], and Washington [Was]. I would like to thank my former students K. Belabas, C. Delaunay, S. Duquesne, and D. Simon, who have helped me to write specific sections, and my colleagues J.-F. Jaulent and J. Martinet for answering many questions in algebraic number theory. I would also like to thank M. Bennett, J. Cremona, A. Kraus, and F. Rodriguez-Villegas for valuable comments on parts of this book. I want especially like to thank D. Bernardi for his thorough rereading of the first ten chapters of the manuscript, which enabled me to remove a large number of errors, mathematical or otherwise.

It is unavoidable that there still remain errors, typographical or otherwise, and the author would like to hear about them. Please send e-mail to

`Henri.Cohen@math.u-bordeaux1.fr`

Lists of known errors for the author's books including the present one can be obtained on the author's home page at the URL

`http://www.math.u-bordeaux1.fr/~cohen/`

Table of Contents

Part II. Diophantine Equations

Part III. Analytic Methods

Part IV. Modern Methods