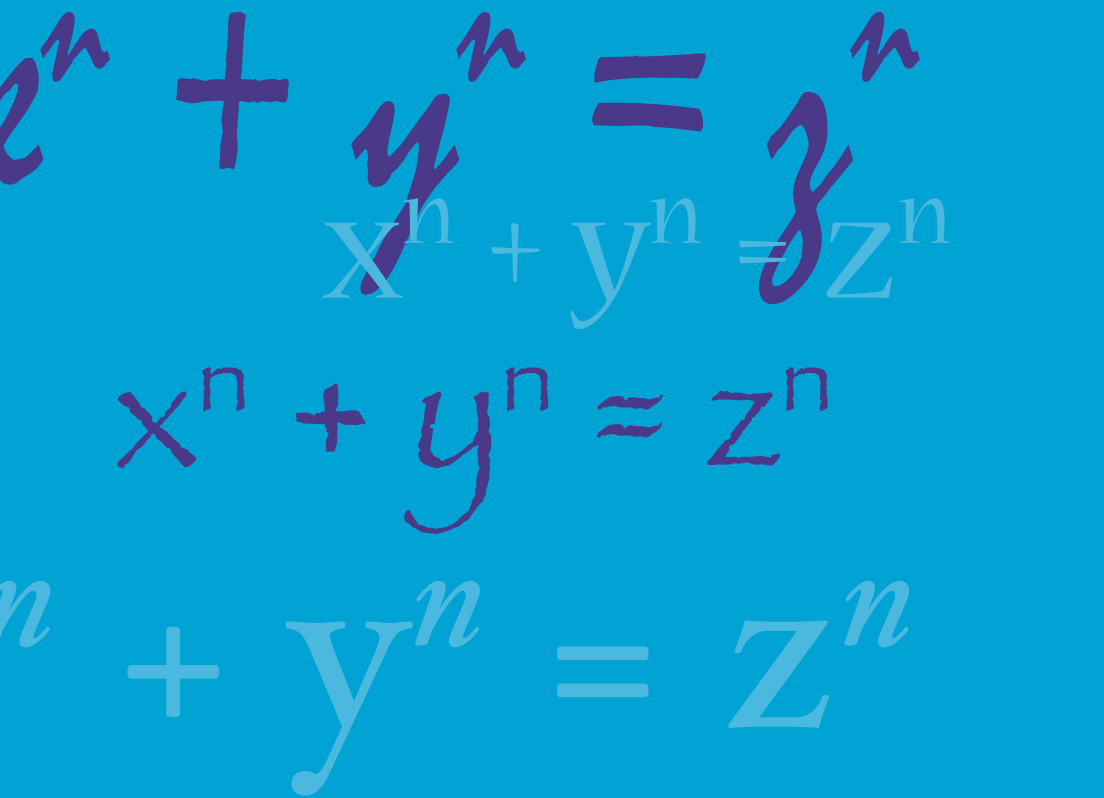
Algebraic Number Theory and Fermat's Last Theorem

FOURTH EDITION



Ian Stewart • David Tall

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Algebraic Number Theory and Fermat's Last Theorem

FOURTH EDITION

This page intentionally left blank

Algebraic Number Theory and Fermat's Last Theorem

FOURTH EDITION

Ian Stewart
University of Warwick
United Kingdom

David Tall
University of Warwick
United Kingdom

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20150616

International Standard Book Number-13: 978-1-4987-3840-8 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface to the Third Edition

The title of this book indicates a dual purpose. Our first aim is to introduce fundamental ideas of algebraic numbers. The second is to tell one of the most intriguing stories in the history of mathematics—the quest for a proof of Fermat's Last Theorem. We use this celebrated theorem to motivate a general study of the theory of algebraic numbers, from a reasonably concrete point of view. The range of topics that we cover is selected to allow students to make early progress in understanding the necessary concepts.

'Algebraic Number Theory' can be read in two distinct ways. One is the theory of numbers viewed algebraically, the other is the study of algebraic numbers. Both apply here. We illustrate how basic notions from the theory of algebraic numbers may be used to solve problems in number theory. However, our main focus is to extend properties of the natural numbers to more general number structures: algebraic number fields, and their rings of algebraic integers. These structures have most of the standard properties that we associate with ordinary whole numbers, but some subtle properties concerning primes and factorization sometimes fail to generalize.

A Diophantine equation (named after Diophantus of Alexandria, who—it is thought—lived around 250 and whose book *Arithmetica* systematized such concepts) is a polynomial equation, or a system of polynomial equations, that is to be solved in integers or rational numbers. The central problem of this book concerns solutions of a very special Diophantine equation:

$$x^n + y^n = z^n$$

where the exponent n is a positive integer. For $n = 2$ there are many integer solutions—in fact, infinitely many—which neatly relate to the theorem of

Pythagoras. For $n \geq 3$, however, there appear to be no integer solutions. It is this assertion that became known as Fermat's Last Theorem. (It is equivalent to there being no rational solutions—try to work out why.)

One method of attack might be to imagine the equation $x^n + y^n = z^n$ as being situated in the complex numbers, and to use the complex nth root of unity $\zeta = e^{2\pi i/n}$ to obtain the factorization (valid for odd n)

$$x^n + y^n = (x + y)(x + \zeta y) \dots (x + \zeta^{n-1} y).$$

This approach entails introducing algebraic ideas, including the notion of factorization in the ring $\mathbf{Z}[\zeta]$ of polynomials in ζ. This promising line of attack was pursued for a time in the 19th century, until it was discovered that this particular ring of algebraic numbers does not possess all of the properties that it 'ought to'. In particular, factorization into 'primes' is not unique in this ring. (It fails, for instance, when $n = 23$, although this is not entirely obvious.) It took a while for this idea to be fully understood and for its consequences to sink in, but as it did so, the theory of algebraic numbers was developed and refined, leading to substantial improvements in our knowledge of Diophantine equations. In particular, it became possible to prove Fermat's Last Theorem in a whole range of special cases. Subsequently, geometric methods and other approaches were introduced to make further gains, until, at the end of the 20th century, Andrew Wiles finally set the last links in place to establish the proof after a three hundred year search.

To gain insight into this extended story we must assume a certain level of algebraic background. Our choice is to start with fundamental ideas that are usually introduced into algebra courses, such as commutative rings, groups and modules. These concepts smooth the way for the modern reader, but they were not explicitly available to the pioneers of the theory. The leading mathematicians in the 19th and early 20th centuries developed and used most of the basic results and techniques of linear algebra—for perhaps a hundred years—without ever defining an abstract vector space. There is no evidence that they suffered as a consequence of this lack of an explicit theory. This historical fact indicates that abstraction can be built only on an already existing body of specific concepts and relationships. This indicates that students will profit from direct contact with the manipulation of examples of number-theoretic concepts, so the text is interspersed with such examples. The algebra that we introduce—which is what we consider necessary for grasping the essentials of the struggle to prove Fermat's Last Theorem—is therefore not as 'abstract' as it might be. We believe that in mathematics it is important to 'get your hands dirty'. This requires struggling with calculations in specific contexts, where the elegance of polished theory may disguise the essential nature of the math-

ematics. For instance, factorization into primes in specific number fields displays the tendency of mathematical objects to take on a life of their own. In some situations something works, in others it does not, and the reasons why are often far from obvious. Without experiencing the struggle in person, it is quite impossible to understand why the pioneers in algebraic number theory had such difficulties. Of such frustrating yet stimulating stuff is the mathematical fabric woven.

We therefore do not begin with later theories that have proved to be of value in a wider range of problems, such as Galois theory, valuation rings, Dedekind domains, and the like. Our purpose is to get students involved in performing calculations that will enable them to build a platform for understanding the theory. However, *some* algebraic background is necessary. We assume a working knowledge of a variety of topics from algebra, reviewed in detail in Chapter 1. These include commutative rings and fields, ideals and quotient rings, factorization of polynomials with real coefficients, field extensions, symmetric polynomials, modules, and free abelian groups. Apart from these concepts we assume only some elementary results from the theory of numbers and a superficial comprehension of multiple integrals.

For organizational reasons rather than mathematical necessity, the book is divided into four parts. Part I develops the basic theory from an algebraic standpoint, introducing the ring of integers of a number field and exploring factorization within it. Quadratic and cyclotomic fields are investigated in more detail, and the Euclidean imaginary fields are classified. We then consider the notion of factorization and see how the notion of a 'prime' p can be pulled apart into two distinct ideas. The first is the concept of being 'irreducible' in the sense that p has no factors other than 1 and p. The second is what we now call 'prime': that if p is a factor of the product ab (possibly multiplied by units—invertible elements) then it must be a factor of either a or b. In this sense, a prime must be irreducible, but an irreducible need not be prime. It turns out that factorization into irreducibles is not always unique in a number field, but useful sufficient conditions for uniqueness may be found. The factorization theory of ideals in a ring of algebraic integers is more satisfactory, in that every ideal is a unique product of prime ideals. The extent to which factorization is not unique can be 'measured' by the group of ideal classes (fractional ideals modulo principal ones).

Part II emphasizes the power of geometric methods arising from Minkowski's theorem on convex sets relative to a lattice. We prove this key result geometrically by looking at the torus that appears as a quotient of Euclidean space by the lattice concerned. As illustrations of these ideas we prove the two- and four-squares theorems of classical number theory; as the main application we prove the finiteness of the class group.

Part III concentrates on applications of the theory thus far developed, beginning with some slightly *ad hoc* computational techniques for class numbers, and leading up to a special case of Fermat's Last Theorem that exemplifies the development of the theory by Kummer, prior to the final push by Wiles.

Part IV describes the final breakthrough, when—after a long period of solitary thinking—Wiles finally put together his proof of Fermat's Last Theorem. Even this tale is not without incident. His first announcement in a lecture series in Cambridge turned out to contain a subtle unproved assumption, and it took another year to rectify the error. However, the proof is finally in a form that has been widely accepted by the mathematical community. In this text we cannot give the full proof in all its glory. Instead we discuss the new ingredients that make the proof possible: the ideas of elliptic curves and elliptic integrals, and the link that shows that the existence of a counterexample to Fermat's Last Theorem would lead to a mathematical construction involving elliptic integrals. The proof of the theorem rests upon showing that such a construction cannot exist. We end with a brief survey of later developments, new conjectures, and open problems.

There follow two appendices which are of importance in algebraic number theory, but do not contribute directly to the proof of Fermat's Last Theorem. The first deals with quadratic residues and the quadratic reciprocity theorem of Gauss. It uses straightforward computational techniques (deceptively so: the ideas are very clever). It may be read at an early stage—for example, right at the beginning, or alongside Chapter 3 which is rather short: the two together would provide a block of work comparable to the remaining chapters in the first part of the book. The second appendix proves the Dirichlet Units Theorem, again a beacon in the development of algebraic number theory, but not directly required in the proof of Fermat's Last Theorem.

A preliminary version of Parts I–III of the book was written in 1974 by Ian Stewart at the University of Tübingen, under the auspices of the Alexander von Humbolt Foundation. This version was used as the basis of a course for students in Warwick in 1975; it was then revised in the light of that experience, and was published by Chapman and Hall. That edition also benefited from the subtle comments of a perceptive but anonymous referee; from the admirable persistence of students attending the course; and from discussions with colleagues. The book has been used by successive generations of students, and a second edition in 1986 brought the story up to date—at that time—and corrected typographical and computational errors.

In the 1980s a proof of Fermat's Last Theorem had not been found. In fact, graffiti on the wall of the Warwick Mathematics Institute declared 'I have a proof that Fermat's Last Theorem is equivalent to The Four Colour Theorem, but this wall is too small for me to write it.' Since that time, both Fermat's Last Theorem and the Four Colour Theorem have fallen, after centuries of effort by the mathematical community. The final conquest of Fermat's Last Theorem required a new version that would give a reasonable idea of the story behind the complete saga. This new version, brought out with a new publisher, is the result of further work to bring the book up to date for the 21st century. It involved substantial rewriting of much of the material, and two new chapters on elliptic curves and elliptic functions. These topics, not touched upon in previous editions, were required to complete the final solution of the most elusive conundrum in pure mathematics of the last three hundred years.

Coventry, February 2001. Ian Stewart
David Tall

This page intentionally left blank

Preface to the Fourth Edition

There are three main changes to this fourth edition.

We provide up-to-date information on what is known about unique prime factorization for real quadratic number fields, especially Malcolm Harper's proof that $\mathbb{Z}(\sqrt{14})$ is Euclidean.

We have added one very important new result: Preda Mihăilescu's stunning proof of the Catalan Conjecture of 1844. This states the only non-trivial consecutive integer powers are $8\ (= 2^3)$ and $9\ (= 3^2)$. We discuss the history of this problem and sketch the current version of the proof, which is an extensive technical application of cyclotomic integers $\mathbb{Z}(\zeta)$ where ζ is a complex root of unity.

Chapter 14 of the previous edition has been split into two separate chapters for reasons of length. Chapter 14 now covers classical ideas about modular functions. Chapter 15 sketches the new ideas of Frey, Wiles and others that led to the long-sought proof of Fermat's Last Theorem. Section 15.4 on recent developments has been updated.

We have also corrected known typographical errors, extended and corrected the index, improved several figures, updated the bibliography and the further reading list, clarified a few historical remarks, and made many small stylistic changes, usually to conform to current practice. Among them is the replacement of boldface symbols such as $\mathbf{R}$ by 'blackboard bold' symbols $\mathbb{R}$.

Coventry and Kenilworth, May 2015.

Ian Stewart
David Tall

This page intentionally left blank

Index of Notation

(q/p)	Legendre symbol
$\mathbb{Z}$	Integers
$\mathbb{Q}$	Rationals
$\mathbb{R}$	Reals
$\mathbb{C}$	Complex numbers
$\mathbb{N}$	Natural numbers
$\mathbb{Z}_n$	Integers modulo n
R/I	Quotient ring
$\ker f$	Kernel of f
$\operatorname{im} f$	Image of f
$\langle X \rangle$	Ideal generated by X
$\langle x_1, \ldots, x_n \rangle$	Ideal generated by $x_1, \ldots, x_n$
$R[t]$	Ring of polynomials over R in t
∂p	Degree of polynomial p
$b \mid a$	b divides a
$\mathrm{D}f$	Formal derivative of f
$L : K$	Field extension
$[L : K]$	Degree of field extension
$K(\alpha_1, \ldots, \alpha_n)$	Field obtained by adjoining $\alpha_1, \ldots, \alpha_n$ to K
$R(\alpha_1, \ldots, \alpha_n)$	Ring generated by R and $\alpha_1, \ldots, \alpha_n$
$s_r(t_1, \ldots, t_n)$	rth elementary symmetric polynomial in $t_1, \ldots, t_n$
N/M	Quotient module
$\langle X \rangle_R$	R-submodule generated by X
$\det(A)$	Determinant of A
(a_{ij})	Matrix
$\mathbb{Z}^n$	Set of n-tuples with integer entries
$\tilde{A}$	Adjoint of matrix A

$\lvert X\rvert$	Cardinality of set x
$\mathbb{A}$	Algebraic numbers
$f_\alpha(t)$	Field polynomial of α
$p_\alpha(t)$	Minimum polynomial of α
ω	$\frac{1}{2}(-1+i\sqrt{3})$
$\Delta[\alpha_1,\ldots,\alpha_n]$	Discriminant of a basis
$\mathbb{B}$	Algebraic integers
$\mathfrak{O}$	Ring of integers of number field
$\mathfrak{O}_K$	Ring of integers of number field K
$\mathrm{N}_K(\alpha)$	Norm of α
$\mathrm{T}_K(\alpha)$	Trace of α
$\mathrm{N}(\alpha)$	Norm of α
$\mathrm{T}(\alpha)$	Trace of α
Δ_G	Discriminant of $\alpha_1,\ldots,\alpha_n$ when this generates G
$\binom{j}{i}$	Binomial coefficient
$U(R)$	Groups of units of R
$\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{p}$, etc.	Ideals
$\mathfrak{a}^{-1}$	Inverse of a fractional ideal
$\mathfrak{a}\,\vert\,\mathfrak{b}$	$\mathfrak{a}$ divides $\mathfrak{b}$: equivalently, $\mathfrak{a}\supseteq\mathfrak{b}$
$\mathrm{N}(\mathfrak{a})$	Norm of $\mathfrak{a}$
$B_r[x]$	Closed ball centre x, radius r
$\lVert x-y\rVert$	Distance from x to y in $\mathbb{R}^n$
$\mathbb{S}$	Circle group
$\mathbb{T}^n$	$\mathbb{N}^n/\mathbb{Z}^n$, the n-dimensional torus
$v(X)$	Volume of X
v	Natural homomorphism $\mathbb{R}^N\to\mathbb{T}^n$
$\mathbb{L}^{st}$	$\mathbb{R}^s\times\mathbb{C}^t$
s	Number of real monomorphisms $K\to\mathbb{C}$
t	Half number of complex monomorphisms $K\to\mathbb{C}$
σ	Map $K\to\mathbb{L}^{st}$
$\mathcal{F}$	Group of fractional ideals
$\mathcal{P}$	Group of principal ideals
$\mathcal{H}$	Class-group $\mathcal{F}\,\vert\,\mathcal{P}$
$h(\mathfrak{O})$	Class-number
h	Class-number
$\mathfrak{a}\sim\mathfrak{b}$	Equivalence of fractional ideals modulo principal ideals
$[\mathfrak{a}]$	Equivalence class of $\mathfrak{a}$
Δ	Discriminant of K
M_{st}	Minkowski constant $(\frac{4}{\pi})^t(s+2t)^{-s-2t}(s+2t)!$
$\mathfrak{I}$	Ideal of $\mathbb{Z}(\zeta)$ generated by $1-\zeta$ where $\zeta=e^{2\pi i/p}$

λ	$1-\zeta$
$\bar{z}$	Complex conjugate of z
B_k	kth Bernouilli number
l	Map $\mathbb{L}^{st} \to \mathbb{R}^{s+t}$
U	Group of units of $\mathfrak{O}$
$\phi(x)$	Euler function
$\mathbb{RP}^2$	Real projective plane
$\mathcal{P}$	The plane $\{(x,y,z) : z = 1\}$
$\mathcal{Q}$	The plane $\{(x,y,z) : z = 0\}$
$\mathbb{CP}^2$	Complex projective plane
$\sim$	Equivalence relation for homogeneous coordinates
g_2, g_3	Coefficients in Weierstrass normal form of a cubic
$\mathbb{O}$	Specific rational point on an elliptic curve
$\mathcal{G}$	Set of rational points on an elliptic curve
$P * Q$	Geometric construction on elliptic curve
$P + Q$	Group operation on elliptic curve
$F(k, v)$	Elliptic integral of the first kind
$\operatorname{sn} u$	Elliptic function
$\operatorname{cn} u$	Elliptic function
$\operatorname{dn} u$	Elliptic function
ω_1, ω_2	Periods of an elliptic function
$\mathcal{L}_{\omega_1,\omega_2}$	Lattice generated by ω_1, ω_2
$\wp$	Weierstrass $\wp$-function
$P \oplus Q$	Renaming of $P + Q$ for clarity
$\mathbb{C} \cup \{\infty\}$	Riemann sphere
$\mathbb{SL}_2(\mathbb{Z})$	Special linear group
$\mathbb{PSL}_2(\mathbb{Z})$	Projective special linear group
$\mathbb{H}$	Upper half-plane in $\mathbb{C}$
$\mathbb{D}$	Modular domain
$X_0(N)$	Modular curve of level N
$\mathcal{F}$	Frey elliptic curve
$P(N)$	Power function of N
$P(A, B, C)$	Power of (A, B, C)

This page intentionally left blank

The Origins of Algebraic Number Theory

Numbers have fascinated the human race for millennia. The Pythagoreans studied many properties of the natural numbers $1, 2, 3, \ldots$, and the famous theorem of Pythagoras, though geometrical, has a pronounced number-theoretic content. Earlier Babylonian civilizations had noted empirically many so-called Pythagorean triples, such as 3, 4, 5 and 5, 12, 13. These are natural numbers a, b, c such that

$$a^2 + b^2 = c^2. \tag{1}$$

A clay tablet from about 1500 BC includes the triple 4961, 6480, 8161, demonstrating the sophisticated techniques of the Babylonians.

The Ancient Greeks, though concentrating on geometry, continued to take an interest in numbers. Around 250 AD, Diophantus of Alexandria wrote a highly influential treatise on polynomial equations which studied solutions in fractions. Particular cases of these equations with natural number solutions have been called *Diophantine* equations to this day.

The study of algebra developed over the centuries, too. Indian and Chinese mathematicians dealt with increasing confidence with negative numbers and zero. Meanwhile the Rashidun Caliphate conquered Alexandria in the 7th century, sweeping across north Africa and Spain. The ensuing civilization brought an enrichment of mathematics with Muslim ingenuity grafted on to Greek and Hindu influence. The word 'algebra' itself derives from the Arabic title 'al jabr w'al muqābalah' (literally 'restoration and equivalence') of a book written by the Persian Al-Khowarizmi in about 825. By the 13th century, peaceful coexistence of Islam and Christianity led to most Greek and Arabic classics being available in Latin translations.

In the 16th century, Girolamo Cardano used negative and imaginary solutions in his famous book *Ars Magna* (The Great Art), and in succeeding centuries complex numbers were used with greater understanding and flexibility.

Meanwhile the theory of natural numbers was not neglected. One of the greatest number theorists of the 17th century was Pierre de Fermat (1601–1665). His fame rests on his correspondence with other mathematicians, for he published very little. He would set challenges in number theory based on his own calculations; and at his death he left a number of theorems whose proofs were known, if at all, only to himself. The most notorious of these was a marginal note in his own personal copy of Diophantus, written in Latin, which translates:

> To resolve a cube into [the sum of] two cubes, a fourth power into fourth powers, or in general any power higher than the second into two of the same kind, is impossible; of which fact I have found a remarkable proof. The margin is too small to contain it.

More precisely, Fermat asserted that, in contrast to the case of Pythagorean triples, the equation

$$x^n + y^n = z^n \tag{2}$$

has no integer solutions x, y, z (other than the trivial ones with one or more of x, y, z equal to zero).

In the years following Fermat's death, almost all of his stated results were furnished with a proof. An exception was his claim that $F_n = 2^{2^n} + 1$ is prime for all positive integers n. In a letter to Pierre de Carcavi in 1659 he claimed a proof of this conjecture, but it was subsequently shown that he was wrong: for instance, F_5 is divisible by 641. Even the great Fermat could make mistakes. But one by one, his other assertions were furnished with proofs, until by the mid-19th century only one elusive jewel remained. A proof of his statement about the non-existence of solutions of (2) for $n \geq 3$ exceeded the powers of all 19th century mathematicians. This beguiling and infuriating assertion, so simple to state, yet so subtle in its labyrinthine complexity, became known as 'Fermat's Last Theorem'. This romantic epithet is in fact doubly inappropriate for, without a proof, it was not a 'theorem', neither was it the last result that Fermat studied—only the last to remain unproved by other mathematicians.

Given that a proof is so elusive, is it really credible that Fermat could have possessed a genuine proof—a clever way of looking at the problem, which eluded later generations? Or had he made a subtle error, which passed unnoticed, so that his 'theorem' had no proof at all? No one knows

for sure, but there is a strong consensus that if he did have what he thought was a proof, it would not survive modern scrutiny. Consensus and certainty are not the same thing, however.

Be that as it may, during the late 19th and early 20th centuries the name stuck, with its glow of romanticism—somehow lacking in the more appropriate title 'Fermat Conjecture'. It has the two classic ingredients of a problem that can capture the imagination of a wider public—a simple statement that can be widely understood, but a proof that defeats the greatest intellects.

Another classic problem of this type—the impossibility of trisecting an angle using only ruler and compasses—took two thousand years to be solved. This problem was posed by the Greeks in their study of geometry; it was solved in the early 19th century using algebraic techniques. In the same way the advancement in the solution of Fermat's Last Theorem has moved away from the original domain, the theory of natural numbers, to a different area of mathematical study, algebraic numbers. By the 19th century the developing theory of algebra had matured to a state where it could usefully be applied in number theory.

As it happened, Fermat's Last Theorem was not the main problem being attacked by number theorists at the time; for example, when Kummer made the all-important breakthrough that we are to describe in this text, he was working on a different problem: 'higher reciprocity laws'. At this stage it is worth making a minor diversion to look at this subject, for it was here that algebraic numbers entered number theory in the work of Carl Friedrich Gauss. In 1796 the eighteen-year-old Gauss had given the first proof of a remarkable fact observed empirically by Leonhard Euler in 1783. Euler had investigated when an integer q is congruent to a perfect square modulo a prime p,

$$x^2 \equiv q \pmod{p}.$$

If so, q is a *quadratic residue* of p. Euler concentrated on the case when p, q are distinct odd primes and noted: if at least one of the odd primes p, q is of the form $4r + 1$, then q is a quadratic residue of p if and only if p is a quadratic residue of q; on the other hand, if both p, q are of the form $4r + 3$, then precisely one is a quadratic residue of the other. However, he failed to find a proof.

Because of the reciprocal nature of the relationship between p and q, this result was known as the *quadratic reciprocity law.* Adrien-Marie Legendre attempted a proof in 1785 but assumed that certain arithmetic series contain infinitely many primes—a theorem whose proof turned out to be far deeper than the quadratic reciprocity law itself. Legendre also introduced

the symbol

$$(q/p) = \begin{cases} 1 & \text{if } q \text{ is a quadratic residue of } p \\ -1 & \text{if not,} \end{cases}$$

in terms of which the law becomes

$$(q/p)(p/q) = (-1)^{(p-1)(q-1)/4}.$$

We now call this the Legendre symbol. It is commonly written

$$\left(\frac{q}{p}\right)$$

but (q/p) is more convenient typographically.

Gauss gave the first proof of the law of quadratic reciprocity in 1796, but he was dissatisfied because his method did not seem a natural way to attack so seemingly simple a theorem. He went on to give several more proofs, two of which appeared in his book *Disquisitiones Arithmeticae* (1801), a definitive text on number theory which still remains in print, Gauss [32]. His second proof depends on a numerical criterion that he discovered, and we give a computational proof depending on this criterion in Appendix A.

Between 1808 and 1832 Gauss continued to look for similar laws for powers higher than squares. This entailed looking for relationships between p and q so that q is a cubic residue of p ($x^3 \equiv q \pmod{p}$) or a biquadratic residue ($x^4 \equiv q \pmod{p}$), and so on. He found some partial results about higher reciprocity laws, and in doing so he discovered that his calculations were simplified by working over the Gaussian integers $a + bi$ ($a, b \in \mathbb{Z}, i = \sqrt{-1}$), rather than the integers alone. This led him to develop a theory of prime factorization for Gaussian integers. He proved that decomposition into primes is unique in that context, and from that he developed a law of biquadratic reciprocity. In the same way, he considered cubic reciprocity by using numbers of the form $a + b\omega$ where $\omega = e^{(2\pi i)/3}$. These higher reciprocity laws do not have the same striking simplicity as quadratic reciprocity, and we shall not study them in this text. But Gauss's use of these new types of number is of fundamental importance for Fermat's Last Theorem, and the study of their factorization properties is a deep and fruitful source of methods and problems.

The numbers concerned are all examples of a particular type of complex number, namely one that is a solution of a polynomial equation

$$a_n x^n + \ldots + a_1 x + a_0 = 0$$

where all the coefficients a_j are integers. Such a complex number is said to be *algebraic*; if further $a_n = 1$ it is called an *algebraic integer*. Examples

of algebraic integers include i (which satisfies $x^2+1=0$), $\sqrt{2}\,(x^2-2=0)$ and more complicated examples, such as the roots of $x^7-265x^3+7x^2-2x+329=0$. The number $\frac{1}{2}i$ (satisfying $4x^2+1=0$) is algebraic but not an integer. On the other hand, there are complex numbers which are not algebraic, such as e or π, although proofs of those statements are difficult.

In the wider setting of algebraic integers, we can factorize a solution of Fermat's equation $x^n+y^n=z^n$ (if one exists) by introducing a complex nth root of unity $\zeta=e^{2\pi i/n}$ and writing (2) as

$$(x+y)(x+\zeta y)\dots(x+\zeta^{n-1}y)=z^n. \tag{3}$$

If $\mathbb{Z}[\zeta]$ denotes the set of algebraic integers of the form $a_0+a_1\zeta+\ldots+a_r\zeta^r$ where each a_r is an ordinary integer, then this factorization takes place in the ring $\mathbb{Z}[\zeta]$.

In 1847 the French mathematician Gabriel Lamé announced a proof of Fermat's Last Theorem. In outline his proposal was to show that only the case where x, y have no common factors need be considered, and then deduce that in this case $x+y, x+\zeta y, \dots, x+\zeta^{n-1}y$ have no common factors, that is, they are relatively prime. He then argued that a product of relatively prime numbers in (3) can equal an nth power only if each factor is an nth power. So

$$\begin{aligned} x+y &= u_1^n \\ x+\zeta y &= u_2^n \\ &\vdots \\ x+\zeta^{n-1} &= u_n^n \end{aligned} \tag{4}$$

On this basis Lamé derived a contradiction.

Announcing a proof does not imply that it is one. Joseph Liouville immediately pointed out that the deduction of (4) from (3) assumes uniqueness of factorization in a subtle manner. Liouville's fears were confirmed when he later received a letter from Ernst Kummer, who had shown that uniqueness of factorization fails in some cases, the first being $n=23$. Over the summer of 1847 Kummer went on to devise his own proof of Fermat's Last Theorem for certain exponents n, surmounting the difficulties of non-uniqueness of factorization by introducing the theory of 'ideal' complex numbers. In retrospect this theory can be viewed as introducing numbers from outside $\mathbb{Z}[\zeta]$ to use as factors when factorizing elements within $\mathbb{Z}[\zeta]$. These 'ideal factors' restore a version of unique factorization.

Subsequently the theory began to take on a different form from that in which Kummer had left it, but the key concept of an 'ideal'—a reformulation by Richard Dedekind of Kummer's 'ideal number'—gave the theory a major boost. By using his theory of ideal numbers, Kummer proved

Fermat's Last Theorem for a wide range of prime powers—the so-called 'regular' primes. He also evolved a powerful machine with applications to many other problems in mathematics. In fact a large part of classical number theory can be expressed in the framework of algebraic numbers. This point of view was urged most strongly by David Hilbert in his *Zahlbericht* (Number Report) of 1897, which had an enormous influence on the development of number theory, see Reid [63].

As a result, algebraic number theory today is a flourishing and important branch of mathematics, with deep methods and insights, and—most significantly—applications not only to number theory, but also to group theory, algebraic geometry, topology, and analysis. It was these wider links that eventually led to the final proof of Fermat's Last Theorem, establishing it once and for all as a theorem, not a conjecture. The eventual proof was made possible by various significant inroads, which were made using techniques from elliptic functions, modular forms, and Galois representations.

The breakthrough, as indicated above, was made by Andrew Wiles. As a teenager, fascinated by the simplicity of the statement of the theorem, Wiles had begun a long and mostly solitary journey in search of a proof. The event that triggered his final push was a conjecture put forward by two Japanese mathematicians, Yutaka Taniyama and Goro Shimura, who hypothesized a link between elliptic curves and modular forms. Their ideas were later refined by André Weil. This proposal became known as the Taniyama–Shimura–Weil Conjecture, and it was discovered that if this conjecture could be proved, then Fermat's Last Theorem could be deduced from it. At this point, Wiles leaped into action. He worked in solitude for seven years before he convinced himself that he had proved a special case of the Taniyama–Shimura–Weil Conjecture that was strong enough to imply Fermat's Last Theorem. He announced his result in a lecture in Cambridge on 23 June 1993.

When his proof was being checked, a query from a colleague revealed a gap, and Wiles accepted that some details required attention. It took him so long to do this that some questioned whether he had ever been close to the proof at all. However, in the autumn of 1994, working with his former student Richard Taylor, he finally realised that he could complete the proof satisfactorily. He released the proof for scrutiny in October 1994 and it was published in May 1995.

Fermat's Last Theorem probably has the distinction of being the theorem with the greatest number of false 'proofs', so the proof was scrutinized very carefully. However, this time the ideas fitted together so tightly that experts in the mathematical community agreed that all was well. In the ensuing period nothing has happened to change this opinion: Fermat's Last Theorem has at last been declared true. However, the proof uses techniques

far beyond what would have been available to Fermat. So when he stated that he had found a proof that could not be fitted into the margin of his book, had he truly found a perceptive insight that has been missed by mathematicians for over 350 years? Or was it, as observed by the historian Dirk Struik [81], that 'even the great Fermat slept sometimes'?

This page intentionally left blank

I

Algebraic Methods

This page intentionally left blank

1

Algebraic Background

Fermat's Last Theorem is a special problem in the general theory of Diophantine equations—integer solutions of polynomial equations. To place the problem in context, we move to the wider realm of algebraic numbers, which arise as the real or complex solutions of polynomials with integer coefficients; we focus particularly on algebraic integers, which are solutions of polynomials with integer coefficients where the leading coefficient is 1. For example, the equation $x^2 - 2 = 0$ has no integer solutions, but it has two real solutions, $x = \pm\sqrt{2}$. The leading coefficient of the polynomial $x^2 - 2$ is 1, so $\pm\sqrt{2}$ are algebraic integers.

To operate with such numbers, it is useful to work in subsystems of the complex numbers that are closed under the usual operations of arithmetic. Such subsystems include subrings (which are closed under addition, subtraction and multiplication) and subfields (closed under all four arithmetic operations including division). Thus along with $\pm\sqrt{2}$ we consider the ring of all numbers $a + b\sqrt{2}$ for $a, b \in \mathbb{Z}$, and the field of all numbers $p + q\sqrt{2}$ for $p, q \in \mathbb{Q}$.

In this chapter we lay the foundations for algebraic number theory by considering some fundamental facts about rings, fields, and other algebraic structures, including abelian groups and modules, which are relevant to our theoretical development. We expect the reader to be acquainted with elementary properties of groups, rings and fields, and to have a basic knowledge of linear algebra over an arbitrary field, up to simple properties of determinants. Familiar results at this level will be stated without proof; results that we think may be less familiar to some readers are proved in full or in outline as appropriate. References are given for results not proved in full, in case the reader wishes to pursue them in greater depth. Useful

general references on abstract algebra, with emphasis on rings and fields, are Fenrick [27], Fraleigh [28], Jacobson [44], Lang [47], Sharpe [74], and Stewart and Tall [79]. For group theory, see Burn [12], Humphreys [42], Macdonald [49], Neumann *et al.* [61], and Rotman [69].

First we set up the ring-theoretic language, in particular the notion of an ideal, which proves to be so important. Then we consider factorization of polynomials over a ring, which in this book is often a subring of the complex numbers. Topics of central importance at this stage are factorization of a polynomial over an extension field, and the theory of elementary symmetric polynomials. Module-theoretic language helps us to clarify certain points later. Results concerning finitely generated abelian groups are proved because they are vital in describing the additive group structure of the subrings of the complex numbers that occur.

1.1 Rings and Fields

Unless explicitly stated to the contrary, the term *ring* in this book will always mean a commutative ring R with identity element 1 (or 1_R). If such a ring has no zero-divisors (so that in R, $a \neq 0$, $b \neq 0$ implies $ab \neq 0$), and if $1 \neq 0$ in it, then it is called a *domain*. (Another common term is *integral domain*, but we omit 'integral' throughout.) An element a in a ring R is called a *unit* if there exists $b \in R$ such that $ab = 1$. Suppose $ab = ac = 1$. Then $c = 1c = abc = acb = 1b = b$. The unique b such that $ab = 1$ is denoted by a^{-1}, and ca^{-1} is also denoted by c/a. If $1 \neq 0$ in R and every non-zero element in R is a unit, then R is called a *field*.

We use standard notation $\mathbb{N}$ for the set of natural numbers $0, 1, 2, \ldots$, $\mathbb{Z}$ for the integers, $\mathbb{Q}$ for the rationals, $\mathbb{R}$ for the reals and $\mathbb{C}$ for the complex numbers. Under the usual operations $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields, $\mathbb{Z}$ is a domain, and $\mathbb{N}$ is not even a ring. For $n \in \mathbb{N}$, $n > 0$, we denote the ring of integers modulo n by $\mathbb{Z}_n$. If n is composite, then $\mathbb{Z}_n$ has zero divisors, but for n prime, then $\mathbb{Z}_n$, is a field; see Fraleigh [28] p. 217.

Our convention is that a subring S of a ring R is required to contain 1_R. We can check that S is a subring by demonstrating that $1_R \in S$, and if $s, t \in S$ then $s + t, -s, st \in S$. The subset S then forms a ring in its own right under the operations restricted from R. In the same way, if K is a field, then a subfield F of K is a subset that is a field under the operations restricted from K. We can check that F is a subfield of K by demonstrating that $1_k \in F$, and if $s,\ t \in F \quad (s \neq 0)$ then $s + t$, $-s$, st, $s^{-1} \in F$.

The concept of an *ideal* is of central importance in this text. Recall that an ideal is a non-empty subset I of a ring R such that if $r, s \in I$, then

$r - s \in I$, and if $r \in R$, $s \in I$ then $rs \in I$. We also require the concept of the *quotient ring* R/I of R by an ideal I. The elements of R/I are cosets $I + r$ of the additive group of I in R, with addition and multiplication defined by

$$\begin{aligned}(I + r) + (I + s) &= I + (r + s)\\ (I + r)(I + s) &= I + rs\end{aligned}$$

for all $r, s \in R$. For example, if $n\mathbb{Z}$ is the set of integer multiples of $n \in \mathbb{Z}$, then $\mathbb{Z}/n\mathbb{Z}$ is isomorphic to $\mathbb{Z}_n$.

A *homomorphism* $f : R_1 \to R_2$, where R_1 and R_2 are rings, is a function such that

$$\begin{aligned}f(1_{R_1}) &= 1_{R_2}\\ f(r + s) &= f(r) + f(s)\\ f(rs) &= f(r)f(s)\end{aligned}$$

for all $r, s \in R_1$. A *monomorphism* is an injective (1–1) homomorphism and an *isomorphism* is a bijective (1–1 and onto) homomorphism.

The *kernel* and *image* of a homomorphism f are defined in the usual way:

$$\begin{aligned}\ker f &= \{r \in R_1 \mid f(r) = 0\}\\ \text{im } f &= \{f(r) \in R_2 \mid r \in R_1\}.\end{aligned}$$

The kernel is an ideal of R_1; the image is a subring of R_2; and the *isomorphism theorem* states that there is an isomorphism from $R_1/\ker f$ to $\text{im} f$. For details, see Fraleigh [28], Jacobson [44], or Sharpe [74].

If X and Y are subsets of a ring R we write $X + Y$ for the set of all elements $x + y\,(x \in X, y \in Y)$, and XY for the set of all finite sums $\Sigma x_i y_i\,(x_i \in X, y_i \in Y)$. When X and Y are both ideals, so are $X + Y$ and XY.

The sum $X+Y$ of two subsets can be generalized to an arbitrary collection $\{X_i\}_{i\in I}$ by defining $\Sigma_{i\in I} X_i$ to be the set of all finite sums $x_{i_1}+\ldots+x_{i_n}$ of elements $x_{i_j} \in X_{i_j}$.

We make the customary compression of notation with regard to $\{x\}$ and x, writing for example xY for $\{x\}Y$, $x+Y$ for $\{x\}+Y$, and 0 for $\{0\}$.

The ideal *generated* by a subset X of R is the smallest ideal of R containing X; we denote this by $\langle X\rangle$. If $X = \{x_1, \ldots, x_n\}$, we write $\langle X\rangle$ as $\langle x_1, \ldots, x_n\rangle$. (Some writers use (X) where we have written $\langle X\rangle$, but then the last-mentioned simplification of notation would reduce to the notation for an n-tuple $(x_1, \ldots, x_n)$, so $\langle X\rangle$ is to be preferred.)