Graduate Texts in Mathematics 228

More information about this series at http://www.springer.com/series/136

Graduate Texts in Mathematics

(continued after references)

Fred Diamond
Jerry Shurman

A First Course in Modular Forms

Springer

Fred Diamond
Department of Mathematics
King's College London
Strand, London WC2R 2LS
United Kingdom
fred.diamond@kcl.ac.uk

Jerry Shurman
Department of Mathematics
Reed College
Portland, OR 97202
USA
jerry@reed.edu

Mathematics Subject Classification (2000): 25001, 11019

Library of Congress Cataloging-in-Publication Data
Diamond, Fred.
A first course in modular forms / Fred Diamond and Jerry Shurman.
p. cm. — (Graduate texts in mathematics ; 228)
Includes bibliographical references and index.
ISBN 0-387-23229-X
1. Forms, Modular. I. Shurman, Jerry Michael. II. Title. III. Series.
QA243.D47 2005
512.7′3—dc22 2004058971

ISBN 0-387-23229-X
ISBN 978-0-387-23229-4 (hardcover)
ISBN 978-1-4419-2005-8 (softcover)
DOI 10.1007/978-0-387-27226-9

Printed on acid-free paper.
ISBN 978-0-387-27226-9 (eBook)

Printed in the United States of America.

9 8 7 6 5 4

Springer is part of Springer Science+Business Media (www.springer.com)

For our parents

Contents

Preface

This book explains a result called the Modularity Theorem:

All rational elliptic curves arise from modular forms.

Taniyama first suggested in the 1950's that a statement along these lines might be true, and a precise conjecture was formulated by Shimura. A paper of Weil [Wei67] provided strong theoretical evidence for the conjecture. The theorem was proved for a large class of elliptic curves by Wiles [Wil95] with a key ingredient supplied by joint work with Taylor [TW95], completing the proof of Fermat's Last Theorem after some 350 years. The Modularity Theorem was proved completely by Breuil, Conrad, Taylor, and the first author of this book [BCDT01]. Different forms of it are stated here in Chapters 2, 6, 7, 8, and 9.

To describe the theorem very simply for now, first consider a situation from elementary number theory. Take a quadratic equation

$$Q : x^2 = d, \qquad d \in \mathbb{Z},\ d \neq 0,$$

and for each prime number p define an integer $a_p(Q)$,

$$a_p(Q) = \begin{pmatrix} \text{the number of solutions } x \text{ of equation } Q \\ \text{working modulo } p \end{pmatrix} - 1.$$

The values $a_p(Q)$ extend multiplicatively to values $a_n(Q)$ for all positive integers n, meaning that $a_{mn}(Q) = a_m(Q)a_n(Q)$ for all m and n.

Since by definition $a_p(Q)$ is the Legendre symbol (d/p) for all $p > 2$, one statement of the Quadratic Reciprocity Theorem is that $a_p(Q)$ depends only on the value of p modulo $4|d|$. This can be reinterpreted as a statement that the sequence of solution-counts $\{a_2(Q), a_3(Q), a_5(Q), \dots\}$ arises as a system of eigenvalues on a finite-dimensional complex vector space associated to the equation Q. Let $N = 4|d|$, let $G = (\mathbb{Z}/N\mathbb{Z})^*$ be the multiplicative group of integer residue classes modulo N, and let V_N be the vector space of complex-valued functions on G,

$$V_N = \{f : G \longrightarrow \mathbb{C}\}.$$

For each prime p define a linear operator T_p on V_N,

$$T_p : V_N \longrightarrow V_N, \qquad (T_p f)(n) = \begin{cases} f(pn) & \text{if } p \nmid N, \\ 0 & \text{if } p \mid N, \end{cases}$$

where the product $pn \in G$ uses the reduction of p modulo N. Consider a particular function $f = f_Q$ in V_N,

$$f : G \longrightarrow \mathbb{C}, \qquad f(n) = a_n(Q) \text{ for } n \in G.$$

This is well defined by Quadratic Reciprocity as stated above. It is immediate that f is an eigenvector for the operators T_p,

$$\begin{aligned}(T_p f)(n) &= \begin{cases} f(pn) = a_{pn}(Q) = a_p(Q)a_n(Q) & \text{if } p \nmid N, \\ 0 & \text{if } p \mid N \end{cases} \\ &= a_p(Q)f(n) \quad \text{in all cases.}\end{aligned}$$

That is,

$$T_p f = a_p(Q) f \quad \text{for all primes } p.$$

This shows that the sequence $\{a_p(Q)\}$ is a system of eigenvalues as claimed.

The Modularity Theorem can be viewed as giving an analogous result. Consider a cubic equation

$$E : y^2 = 4x^3 - g_2 x - g_3, \qquad g_2, g_3 \in \mathbb{Z},\ g_2^3 - 27g_3^2 \neq 0.$$

Such equations define *elliptic curves*, objects central to this book. For each prime number p define a number $a_p(E)$ akin to $a_p(Q)$ from before,

$$a_p(E) = p - \begin{pmatrix} \text{the number of solutions } (x, y) \text{ of equation } E \\ \text{working modulo } p \end{pmatrix}.$$

One statement of Modularity is that again the sequence of solution-counts $\{a_p(E)\}$ arises as a system of eigenvalues. Understanding this requires some vocabulary.

A *modular form* is a function on the complex upper half plane that satisfies certain transformation conditions and holomorphy conditions. Let τ be a variable in the upper half plane. Then a modular form necessarily has a Fourier expansion,

$$f(\tau) = \sum_{n=0}^{\infty} a_n(f) e^{2\pi i n \tau}, \quad a_n(f) \in \mathbb{C} \text{ for all } n.$$

Each nonzero modular form has two associated integers k and N called its *weight* and its *level*. The modular forms of any given weight and level form a vector space. Linear operators called the *Hecke operators*, including an operator T_p for each prime p, act on these vector spaces. An *eigenform* is a modular form that is a simultaneous eigenvector for all the Hecke operators. By analogy to the situation from elementary number theory, the Modularity Theorem associates to the equation E an eigenform $f = f_E$ in a vector space V_N of weight 2 modular forms at a level N called the *conductor* of E. The eigenvalues of f are its Fourier coefficients,

$$T_p f = a_p(f)f \quad \text{for all primes } p,$$

and a version of Modularity is that *the Fourier coefficients give the solution-counts*,

$$a_p(f) = a_p(E) \quad \text{for all primes } p. \tag{0.1}$$

That is, the solution-counts of equation E are a system of eigenvalues, like the solution-counts of equation Q, but this time they arise from modular forms,

$$T_p f = a_p(E)f \quad \text{for all primes } p.$$

This version of the Modularity Theorem will be stated in Chapter 8.

Chapter 1 gives the basic definitions and some first examples of modular forms. It introduces elliptic curves in the context of the complex numbers, where they are defined as tori and then related to equations like E but with $g_2, g_3 \in \mathbb{C}$. And it introduces *modular curves*, quotients of the upper half plane that are in some sense more natural domains of modular forms than the upper half plane itself. Complex elliptic curves are compact Riemann surfaces, meaning they are indistinguishable in the small from the complex plane. Chapter 2 shows that modular curves can be made into compact Riemann surfaces as well. It ends with the book's first statement of the Modularity Theorem, relating elliptic curves and modular curves as Riemann surfaces: *If the complex number $j = 1728g_2^3/(g_2^3 - 27g_3^2)$ is rational then the elliptic curve is the holomorphic image of a modular curve.* This is notated

$$X_0(N) \longrightarrow E.$$

Much of what follows over the next six chapters is carried out with an eye to going from this complex analytic version of Modularity to the arithmetic version (0.1). Thus this book's aim is not to prove Modularity but to state its different versions, showing some of the relations among them and how they connect to different areas of mathematics.

Modular forms make up finite-dimensional vector spaces. To compute their dimensions Chapter 3 further studies modular curves as Riemann surfaces. Two complementary types of modular forms are *Eisenstein series* and *cusp forms*. Chapter 4 discusses Eisenstein series and computes their Fourier expansions. In the process it introduces ideas that will be used later in the book, especially the idea of an *L-function*,

$$L(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}.$$

Here s is a complex variable restricted to some right half plane to make the series converge, and the coefficients a_n can arise from different contexts. For instance, they can be the Fourier coefficients $a_n(f)$ of a modular form. Chapter 5 shows that if f is a Hecke eigenform of weight 2 and level N then its L-function has an *Euler factorization*

$$L(s, f) = \prod_p (1 - a_p(f)p^{-s} + \mathbf{1}_N(p)p^{1-2s})^{-1}.$$

The product is taken over primes p, and $\mathbf{1}_N(p)$ is 1 when $p \nmid N$ (true for all but finitely many p) but is 0 when $p \mid N$.

Chapter 6 introduces the *Jacobian* of a modular curve, a complex torus like a complex elliptic curve but possibly of higher dimension. The Jacobian thus has Abelian group structure. Another version of the Modularity Theorem says that every complex elliptic curve with a rational j-value is the holomorphic homomorphic image of a Jacobian,

$$\mathrm{J}_0(N) \longrightarrow E.$$

Modularity refines to say that the elliptic curve is in fact the image of a quotient of a Jacobian, the *Abelian variety* associated to a weight 2 eigenform,

$$A_f \longrightarrow E.$$

This version of Modularity associates a cusp form f to the elliptic curve E.

Chapter 7 brings algebraic geometry into the picture and moves toward number theory by shifting the environment from the complex numbers to the rational numbers. Every complex elliptic curve with rational j-invariant can be associated to the solution set of an equation E with $g_2, g_3 \in \mathbb{Q}$. Modular curves, Jacobians, and Abelian varieties are similarly associated to solution sets of systems of polynomial equations over $\mathbb{Q}$, algebraic objects in contrast to the earlier complex analytic ones. The formulations of Modularity already in play rephrase algebraically to statements about objects and maps defined by polynomials over $\mathbb{Q}$,

$$X_0(N)_{\mathrm{alg}} \longrightarrow E, \qquad \mathrm{J}_0(N)_{\mathrm{alg}} \longrightarrow E, \qquad A_{f,\mathrm{alg}} \longrightarrow E.$$

We discuss only the first of these in detail since $X_0(N)_{\mathrm{alg}}$ is a curve while $\mathrm{J}_0(N)_{\mathrm{alg}}$ and $A_{f,\mathrm{alg}}$ are higher-dimensional objects beyond the scope of this book. These algebraic versions of Modularity have applications to number theory, for example constructing rational points on elliptic curves using points called Heegner points on modular curves.

Chapter 8 develops the *Eichler–Shimura relation*, describing the Hecke operator T_p in characteristic p. This relation and the versions of Modularity already stated help to establish two more versions of the Modularity Theorem. One is the arithmetic version that $a_p(f) = a_p(E)$ for all p, as above. For the other, define the *Hasse–Weil L-function* of an elliptic curve E in terms of the solution-counts $a_p(E)$ and the conductor N of E,

$$L(s, E) = \prod_p (1 - a_p(E)p^{-s} + \mathbf{1}_N(p)p^{1-2s})^{-1}.$$

Comparing this to the Euler product form of $L(s, f)$ above gives a version of Modularity equivalent to the arithmetic one: *The L-function of the modular form is the L-function of the elliptic curve*,

$$L(s, f) = L(s, E).$$

As a function of the complex variable s, both L-functions are initially defined only on a right half plane, but Chapter 5 shows that $L(s, f)$ extends analytically to all of $\mathbb{C}$. By Modularity the same now holds for $L(s, E)$. This is important because we want to understand E as an Abelian group, and the conjecture of Birch and Swinnerton-Dyer is that the analytically continued $L(s, E)$ contains information about the group's structure.

Chapter 9 introduces *ℓ-adic Galois representations*, certain homomorphisms of Galois groups into matrix groups. The simplest nontrivial such representation arises from the Quadratic Reciprocity example at the beginning of this preface. Galois representations are also associated to elliptic curves and to modular forms, incorporating the ideas from Chapters 6 through 8 into a framework rich in additional algebraic structure. The corresponding version of the Modularity Theorem is: *Every Galois representation associated to an elliptic curve over $\mathbb{Q}$ arises from a Galois representation associated to a modular form*,

$$\rho_{f,\ell} \sim \rho_{E,\ell}.$$

This is the version of Modularity that was proved. The book ends by discussing the broader relation between Galois representations and modular forms.

Many good books on modular forms already exist, but they can be daunting for a beginner. Although some of the difficulty lies in the material itself, the authors believe that a more expansive narrative with exercises will help students into the subject. We also believe that algebraic aspects of modular forms, necessary to understand their role in number theory, can be made accessible to students without previous background in algebraic number theory and algebraic geometry. In the last four chapters we have tried to do so by introducing elements of these subjects as necessary but not letting them take over the text. We gratefully acknowledge our debt to the other books, especially to Shimura [Shi73].

The minimal prerequisites are undergraduate semester courses in linear algebra, modern algebra, real analysis, complex analysis, and elementary number theory. Topics such as holomorphic and meromorphic functions, congruences, Euler's totient function, the Chinese Remainder Theorem, basics of general point set topology, and the structure theorem for modules over a principal ideal domain are used freely from the beginning, and the Spectral Theorem of linear algebra is cited in Chapter 5. A few facts about representations and tensor products are also cited in Chapter 5, and Galois theory is used extensively in the later chapters. Chapter 3 quotes formulas from Riemann surface theory, and later in the book Chapters 6 through 9 cite steadily more results from Riemann surface theory, algebraic geometry, and algebraic number theory. Seeing these presented in context should help the reader absorb the new language necessary en route to the arithmetic and representation theoretic versions of Modularity.

We thank our colleagues Joe Buhler, David Cox, Paul Garrett, Cris Poor, Richard Taylor, and David Yuen, Reed College students Asher Auel, Rachel Epstein, Harold Gabel, Michael Lieberman, Peter McMahan, and John Saller, and Brandeis University student Makis Dousmanis for looking at drafts. We thank Joe Lipman and Shaul Zemel for their helpful comments on this book's early printings, and the many other readers who have since alerted us to errors. Comments and corrections should be sent to the second author at jerry@reed.edu. Errata lists for this book's successive printings can be found at http://www.reed.edu/~jerry/. However, the fourth printing contains added subject matter in chapter 5 (mostly Section 5.9) and revisions and updates to chapter 9, creating too many changes from the third printing to detail in a list. The first author received support from NSF Grant 0300434 while working on this book.

July 2004, May 2016

Fred Diamond
King's College London
London, UK

Jerry Shurman
Reed College
Portland, OR