Cambridge studies in advanced mathematics • 69

Modular Forms and Galois Cohomology

HARUZO HIDA

Paperback Re-issue

This book provides a comprehensive account of a key (and perhaps the most important) theory on which the Taylor–Wiles proof of Fermat's last theorem is based. The book begins with an overview of the theory of automorphic forms on linear algebraic groups and then covers the basic theory and results on elliptic modular forms, including a substantial simplification of the proof of Taylor–Wiles by Fujiwara and Diamond. It contains a detailed exposition of the representation theory of profinite groups (including deformation theory), as well as the Euler characteristic formulas of Galois cohomology groups. The final chapter presents a proof of a non-abelian class number formula and includes several new results from the author.

The book will be of interest to graduate students and researchers in number theory (including algebraic and analytic number theorists) and arithmetic algebraic geometry.

Cambridge Studies in Advanced Mathematics

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org

ISBN 978-0-521-07208-3

Cover design by James Butler

69

Modular Forms and Galois Cohomology

HIDA

CAMBRIDGE

Modular Forms and Galois Cohomology

Haruzo Hida
UCLA

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/9780521770361

First published 2000
This digitally printed version 2008

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

ISBN 978-0-521-77036-1 hardback
ISBN 978-0-521-07208-3 paperback

CAMBRIDGE STUDIES IN
ADVANCED MATHEMATICS 69

MODULAR FORMS AND GALOIS COHOMOLOGY

Books in this series

1 W.M.L. Holcombe *Algebraic automata theory*
2 K. Petersen *Ergodic theory*
3 P.T. Johnstone *Stone spaces*
4 W.H. Schikhof *Ultrametric calculus*
5 J.-P. Kahane *Some random series of functions, 2nd edition*
6 H. Cohn *Introduction to the construction of class fields*
7 J. Lambek & P.J. Scott *Introduction to higher-order categorical logic*
8 H. Matsumura *Commutative ring theory*
9 C.B. Thomas *Characteristic classes and the cohomology of finite groups*
10 M. Aschbacher *Finite group theory*
11 J.L. Alperin *Local representation theory*
12 P. Koosis *The logarithmic integral I*
13 A. Pietsch *Eigenvalues and S-numbers*
14 S.J. Patterson *An introduction to the theory of the Riemann zeta-function*
15 H.J. Baues *Algebraic homotopy*
16 V.S. Varadarajan *Introduction to harmonic analysis on semisimple Lie groups*
17 W. Dicks & M. Dunwoody *Groups acting on graphs*
18 L.J. Corwin & F.P. Greenleaf *Representations of nilpotent Lie groups and their applications*
19 R. Fritsch & R. Piccinini *Cellular structures in topology*
20 H Klingen *Introductory lectures on Siegel modular forms*
21 P. Koosis *The logarithmic integral II*
22 M.J. Collins *Representations and characters of finite groups*
24 H. Kunita *Stochastic flows and stochastic differential equations*
25 P. Wojtaszczyk *Banach spaces for analysts*
26 J.E. Gilbert & M.A.M. Murray *Clifford algebras and Dirac operators in harmonic analysis*
27 A. Frohlich & M.J. Taylor *Algebraic number theory*
28 K. Goebel & W.A. Kirk *Topics in metric fixed point theory*
29 J.F. Humphreys *Reflection groups and Coxeter groups*
30 D.J. Benson *Representations and cohomology I*
31 D.J. Benson *Representations and cohomology II*
32 C. Allday & V. Puppe *Cohomological methods in transformation groups*
33 C. Soulé et al *Lectures on Arakelov geometry*
34 A. Ambrosetti & G. Prodi *A primer of nonlinear analysis*
35 J. Palis & F. Takens *Hyperbolicity and sensitive chaotic dynamics at homoclinic bifurcations*
36 M. Auslander, I. Reiten & S. Smalo *Representation theory of Artin algebras*
37 Y. Meyer *Wavelets and operators*
38 C. Weibel *An introduction to homological algebra*
39 W. Bruns & J. Herzog *Cohen-Macaulay rings*
40 V. Snaith *Explicit Brauer induction*
41 G. Laumon *Cohomology of Drinfeld modular varieties I*
42 E.B. Davies *Spectral theory and differential operators*
43 J. Diestel, H. Jarchow & A. Tonge *Absolutely summing operators*
44 P. Mattila *Geometry of sets and measures in Euclidean spaces*
45 R. Pinsky *Positive harmonic functions and diffusion*
46 G. Tenenbaum *Introduction to analytic and probabilistic number theory*
47 C. Peskine *An algebraic introduction to complex projective geometry I*
48 Y. Meyer & R. Coifman *Wavelets and operators II*
49 R. Stanley *Enumerative combinatories*
50 I. Porteous *Clifford algebras and the classical groups*
51 M. Audin *Spinning tops*
52 V. Jurdjevic *Geometric control theory*
53 H. Voelklein *Groups as Galois groups*
54 J. Le Potier *Lectures on vector bundles*
55 D. Bump *Automorphic forms*
56 G. Laumon *Cohomology of Drinfeld modular varieties II*
59 P. Taylor *Practical foundations of mathematics*
60 M. Brodmann & R. Sharp *Local cohomology*
64 J. Jost & X. Li-Jost *Calculus of variations*
66 S. Morosawa et al. *Holomorphic dynamics*
68 Ken-iti Sato *Lévy processes and infinitely divisible distributions*
69 H. Hida *Modular forms and Galois cohomology*

Preface

In the past few years (1995–98), I have given several advanced graduate courses at UCLA in order to provide a comprehensive account of the proof by Wiles (and Taylor) of the identification of certain Hecke algebras with universal deformation rings of Galois representations. Assuming a good knowledge of Class field theory, I started with an overview of the theory of automorphic forms on linear algebraic groups, specifically, $GL(n)$ over number fields. Since second year graduate students often lack knowledge of representation theory of profinite groups, necessary to carry out the task, I went on to describe basic representation theory, the theory of pseudo-representations and their deformation. To reach this point, I had already covered almost a one-year course. Then I continued to give a sketch of the rationality and the control theorems of the space of elliptic modular forms, which is the basis of the definition of the Hecke algebra. In the meantime, K. Fujiwara and F. Diamond independently gave, in 1996, a substantial simplification of the proof of Wiles, which I incorporated in my course. After having proved the theorem, assuming many things, I came back to the material I used in the proof, in particular the duality theorems (due to Poitou and Tate) of Galois cohomology groups. Thus the first chapters follow faithfully my series of courses; so, logically the reader might have to jump around between chapters. However, except for the construction of modular Galois representations, which has been described in some literature already, basically all ingredients of the proof of Wiles are at least covered to some extent. The final chapter (Chapter V) is added after finishing the series of courses, in order to give some indication of further study, and this part contains some new results of mine. An outline of the book can be found in Subsection 2.1 in Chapter 1.

Although I have not covered the proof by Wiles of the Shimura–Taniyama conjecture and Fermat's last theorem, I hope that graduate students can, after finishing this book, thoroughly understand Wiles' original paper treating these two profound results. I hope to return to the theory of elliptic curves and modular Galois representations in a book in the near future.

While I was preparing this book, I received help from many people (including my present and former students) who read the manuscript and provided useful advice and corrections on mathematical, linguistic and historical matters. I wish to thank all of them. I would also like to acknowledge partial support from the national science foundation while I was preparing this book.

May 25th, 1999 at Los Angeles,
Haruzo Hida

Contents

1
Overview of Modular Forms

It is difficult to provide a brief summary of techniques used in modern number theory. Traditionally, mathematical research has been classified by the method mathematicians exploit to study their research areas, except possibly for number theory. For example, algebraists study mathematical questions related to abstract algebraic systems in a purely algebraic way (only allowing axioms defining their algebraic systems), differential geometers study manifolds via infinitesimal analysis, and algebraic geometers study geometry of algebraic varieties (and its siblings) via commutative algebras and category theory. There are no central techniques which distinguish number theory from other subjects, or rather, number theorists exploit any techniques available to hand to solve problems specific to number theory. In this sense, number theory is a discipline in mathematics which cannot be classified by methodology from the above traditional viewpoint but is just a web of rather specific problems (or conjectures) tightly and subtly knit to each other. We just study numbers, those simple ones, like integers, rational numbers, algebraic numbers, real and complex numbers and p-adic numbers, and that is it.

What has emerged from our rather long history is that we continue to study at least two aspects of these numbers: the numbers of the base field and the numbers of its extensions. For example, the *quadratic reciprocity law* describes in a simple way how rational primes decompose as a product of prime ideals in a quadratic extension only using data from rational integers. More generally, by class field theory, we know how rational primes decompose in an abelian extension out of the datum from rational numbers. Thus we have two sets of numbers, the first is the numbers of the base field and the other from an extension of the base field. Nowadays, class field theory is often described using transcendental numbers from all possible completions of the base fields,

involving complex, real and p-adic numbers. The *adele ring* $\mathbb{A}$ is just a subring of $(\prod_p \mathbb{Q}_p) \times \mathbb{R}$ generated by p-adic integers (for all primes p), real numbers and rational numbers (even additively):

$$\mathbb{A} = ((\widehat{\mathbb{Z}} \times \mathbb{R}) + \mathbb{Q}) \subset \left(\prod_p \mathbb{Q}_p\right) \times \mathbb{R} \quad \left(\widehat{\mathbb{Z}} = \prod_p \mathbb{Z}_p\right),$$

where we regard $\mathbb{Q} \subset \mathbb{A}$ by the diagonal embedding $\xi \mapsto (\xi, \xi, \ldots, \xi, \ldots) \in \prod_p \mathbb{Q}_p \times \mathbb{R}$. Thus for a given *number field* F (that is, a finite extension of the rational numbers $\mathbb{Q}$), the adele ring $F_{\mathbb{A}} = F \otimes_{\mathbb{Q}} \mathbb{A}$ of F represents all data from the base field. For a given algebraic group G defined over F, which we may think of as just a coherent rule assigning a group $G(A)$ to any F-algebra A, $G(F_{\mathbb{A}})$ is an immediate source of information. For example, $A \mapsto GL_n(A)$, the group of invertible $n \times n$ matrices with coefficients in A, is an algebraic group. Global class field theory is typically described as a canonical exact sequence:

$$1 \to \overline{GL_1(F)C} \to GL_1(F_{\mathbb{A}}) \to \mathrm{Gal}(F^{ab}/F) \to 1$$

for the identity connected component C of $GL_1(F_{\mathbb{A}})$, where '$\overline{X}$' indicates the topological closure of X, and F^{ab}/F is the composite of all Galois extensions M/F (inside an algebraic closure $\overline{F}$ of F) with $\mathrm{Gal}(M/F)$ abelian (such an extension is called an *abelian extension* of F). Thus we have the second set of numbers F^{ab}: those numbers in a Galois extension specific to our choice of the algebraic group $G = GL_1$. In this first example, $G = GL_1$, which is the simplest (and most important) of all abelian algebraic groups. Thus we might call the study of extensions of a base field the *Galois side* of number theory.

The above example tells us that it is important to study the geometry of the homogeneous space $G(F)\backslash G(F_{\mathbb{A}})$. Most geometers, if they are given a topological space, start studying functions on the space, because they know by experience that functions are easier to manipulate and eventually determine the space. We call functions on $G(F)\backslash G(F_{\mathbb{A}})$ *modular forms*. The homogeneous space $G(F)\backslash G(F_{\mathbb{A}})$ often classifies geometric objects, like abelian varieties and motives (as is often the case for a quotient of a big group by a discrete subgroup, because the big group is somehow a (local) transformation group of a collection of geometric objects, and elements of the discrete subgroup give rise to (global) isomorphisms between the objects). For example, when $G = GL_2$, $X = G(\mathbb{Q})\backslash G(\mathbb{A})/G(\widehat{\mathbb{Z}})Z(\mathbb{R})SO_2(\mathbb{R})$ for the maximal connected compact subgroup $SO_2(\mathbb{R}) \subset GL_2(\mathbb{R})$ and the center $Z(\mathbb{R}) \subset G(\mathbb{R})$ classifies isomorphism classes of elliptic curves over

$\mathbb{C}$, and therefore, gives rise to the set of complex points of the (coarse) moduli scheme $\mathbf{P}^1(j)$ (defined over $\mathbb{Q}$) classifying elliptic curves over $\mathbb{Q}$. Because of the classification property of X, we have a canonical algebraic variety $\mathbf{P}^1(j)$ defined over $\mathbb{Q}$ (and actually defined over $\mathbb{Z}$) which gives rise to X. The scheme $\mathbf{P}^1(j)$ is called a *canonical model* of X. This phenomenon that the homogeneous space $G(F)\backslash G(F_{\mathbb{A}})$ classifies some algebro-geometric objects is prevalent in many other cases of different algebraic groups (like symplectic groups $G = Sp(2g)$ and unitary groups $U(m,n)$), and the resulting canonical models are called *Shimura varieties* of PEL-type. In any case, a general homogeneous space $X(U) = GL_2(\mathbb{Q})\backslash GL_2(\mathbb{A})/U \cdot Z(\mathbb{R})SO_2(\mathbb{R})$ for an open subgroup $U \subset GL_2(\hat{\mathbb{Z}})$ classifies elliptic curves with some additional structure (such as a given point of order N) over $\mathbb{Z}$ (see [AME] and [GMF]). Then the canonical model $X(U)$ is called a *modular curve*, because it is a finite covering of $\mathbf{P}^1(j)$ and hence is an algebraic curve. Thus finding an elliptic curve (with a given additional structure) defined over $\mathbb{Q}$ (or $\mathbb{Z}$) is equivalent to finding a rational (or integral) solution to the defining equations of a specific modular curve $X(U)$. In this way, our effort in understanding the homogeneous space $X(U)$ provides us with another number theoretic question: a *Diophantine problem* of the equations of modular curves. This is a typical example in Number theory of where a serious study of one good problem yields another interesting question, making the life of the theory virtually inexhaustible.

An elliptic curve E defined over a number field $\mathbb{Q}$ is a natural source of a Galois representation $\rho_{E,p} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL_2(\mathbb{Z}_p)$ ramifying at p and a finite set S of primes (independent of p). This comes from the fact that the group $E[p^r]$ of p^r-torsion points of an elliptic curve $E_{/\mathbb{Q}}$ is isomorphic to $(\mathbb{Z}/p^r\mathbb{Z})^2$ and that the Galois action on $E[p^r]$ therefore gives rise to a Galois representation $\rho_E \mod p^r : \mathrm{Gal}(\overline{F}/F) \to GL_2(\mathbb{Z}/p^r\mathbb{Z}) \cong Aut(E[p^r])$. This Galois representation has a remarkable property, found by Hasse, that $L_p(X) = \det(1 - \rho_{E,p}(Frob_\ell)X) = 1 - a(\ell)X + \ell X^2$ has rational integral coefficients $a(\ell)$ independent of p for primes $\ell \notin S \cup \{p\}$ (see, for example, [AME] or [GMF]). Here $Frob_\ell$ is the Frobenius element in the Galois group. Then it is traditional to make an Euler product:

$$L(s,E) = \prod_p L_p(p^{-s})^{-1}.$$

This *Hasse–Weil L-function* is absolutely convergent if $\mathrm{Re}(s) > \frac{3}{2}$, and Hasse and Weil conjectured that it should have an analytic continuation to the whole s-plane with a functional equation relating $L(s,E)$ to $L(2-$

s, E). This is a hard question, because $L(s, E)$ is defined in a purely algebraic way, while the conjecture predicts a purely analytic property (typical for Number theoretic questions, as number theory belongs neither to algebra nor to analysis).

Since a modular form f is a function on a topological group $GL_2(\mathbb{A})$, it is natural to make a convolution product with a compactly supported function ϕ on $GL_2(\mathbb{A})$. This operator $f \mapsto \phi * f$ is called a Hecke operator. Sometime in the 1930's, Hecke discovered that the space of holomorphic modular forms on $X(U)$ has a base made of common eigenforms of standard Hecke operators $T(n)$ indexed by positive integers n (see Section 1.2 for a description of $T(n)$). Pick a common eigenform f, and write the eigenvalues for $T(n)$ as $\lambda(T(n))$. Hecke made an L-function: $L(s, \lambda) = \sum_{n=1}^{\infty} \lambda(T(n)) n^{-s}$. This is a (modular) Hecke L-function, which satisfies a functional equation relating $L(s, \lambda)$ to $L(k - s, \lambda)$ for a positive integer k called the *weight* of f. A remarkable fact is that the eigenvalues are algebraic integers in a number field $\mathbb{Q}(\lambda)$ (as implied by Theorem 3.13 in Chapter 3) independent of n. It is not very often but not rare either that $\lambda(T(n)) \in \mathbb{Z}$ for all n when the weight k is 2 (although $\mathbb{Q}$-rational eigenforms become sporadic as k grows); thus, $\mathbb{Q}(\lambda) = \mathbb{Q}$ in such cases. Another remarkable fact is that this L-function has an Euler product: $L(s, f) = \prod_p H_p(p^{-s})^{-1}$ with an Euler factor $H_p(X) = 1 - a(p)X + \psi(p)p^{k-1}X^2$ for the weight $k \geq 1$ and a Dirichlet character ψ, which is called the *'Neben'* character of f by Hecke. Thus when $k = 2$ and $\psi = 1$, the case Hecke called 'Haupt typus' (principal type), the L-function looks like a Hasse–Weil L-function. Since Hecke initiated the study of the modular side (in the non-abelian case), it would be appropriate to call the study of modular forms (or the numbers of the base field) the *Hecke side* of Number theory.

The *Shimura–Taniyama conjecture* states that the Hasse–Weil L-function of every elliptic curve rational over $\mathbb{Q}$ appears as a Hecke L-function of a rational Hecke eigen cusp form, or equivalently, (and more geometrically) that every $\mathbb{Q}$-rational elliptic curve appears as a factor of the jacobian of a modular curve (see [Lg] and [Sh3] for the history of the conjecture, and see also [Sh4] for an account of Shimura's work in the 50's and 60's). As was shown by Shimura ([IAT] Chapter 7), to each Hecke eigen cusp form of weight 2 defined on a modular curve $X(U)$, one can attach a canonical subabelian variety A (or a quotient) of the jacobian of $X(U)$ so that the L-function of A coincides with the Hecke L-function of the cusp form. This fact implies that a Hecke eigen cusp form with eigenvalue $\lambda(T(\ell))$ and with 'Neben' character ψ has a unique