

# Graduate Texts in Mathematics

Fred Diamond  
Jerry Shurman

## A First Course in Modular Forms

模形式基础教程

Springer

世界图书出版公司  
[www.wpcbj.com.cn](http://www.wpcbj.com.cn)

Fred Diamond  
Jerry Shurman

# A First Course in Modular Forms

 Springer

Fred Diamond  
Department of Mathematics  
Brandeis University  
Waltham, MA 02454  
USA  
fdiamond@brandeis.edu

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

*Editorial Board*

S. Axler  
Mathematics Department  
San Francisco State  
University  
San Francisco, CA 94132  
USA  
axler@sfsu.edu

F.W. Gehring  
Mathematics Department  
East Hall  
University of Michigan  
Ann Arbor, MI 48109  
USA  
fgehring@math.lsa.umich.edu

K.A. Ribet  
Mathematics Department  
University of California,  
Berkeley  
Berkeley, CA 94720-3840  
USA  
ribet@math.berkeley.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

I. Forms, Modular. I. Shurman, Jerry Michael. II. Title. III. Series.

QA243.D47 2005

512.7'3—dc22

2004058971

ISBN 0-387-23229-X

© 2005 Springer Science+Business Media, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

This reprint has been authorized by Springer-Verlag (Berlin/Heidelberg/New York) for sale in the People's Republic of China only and not for export therefrom.

springer.com

---

图书在版编目 (C I P) 数据

模形式基础教程: 英文 / (英) 戴梦德 (F. Diamond),  
(美) 谢尔曼 (Shurman, J.) 著. —北京: 世界图书出版  
公司北京公司, 2007. 5  
(数学研究生丛书)  
书名原文: A first Course in Modular Forms  
ISBN 978-7-5062-8300-7

I. 模… II. ①戴…②谢… III. 模型式—研究生—教材—  
英文 IV. 0156

中国版本图书馆CIP数据核字 (2007) 第058790号

---

---

书 名: A First Course in Modular Forms

作 者: F. Diamond, J. Shurman

中 译 名: 模型式基础教程

责任编辑: 焦小浣

---

出 版 者: 世界图书出版公司北京公司

印 刷 者: 北京世图印刷厂

发 行: 世界图书出版公司北京公司 (北京朝内大街 137 号 100010)

联系电话: 010-64015659, 64038348

电子信箱: kjsk @ vip.sina.com

---

开 本: 24 开

印 张: 19

版 次: 2007 年 5 月第 1 次印刷

版权登记: 图字: 01-2007-1489

---

书 号: 978-7-5062-8300-7/O · 604

定 价: 49.00 元

---

世界图书出版公司北京公司已获得 Springer 授权在中国大陆独家重印发行。

---

## 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

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

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

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

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  $G = (\mathbf{Z}/N\mathbf{Z})^*$  be the multiplicative group of

integer residue classes modulo  $N$ , and let  $V_N$  be the vector space of complex-valued functions on  $G$ ,

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

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

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

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

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

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,  $T_p f = a_p(Q)f$  for all  $p$ . 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

$$E : y^2 = 4x^3 - g_2x - g_3, \quad g_2, g_3 \in \mathbb{Z}, \quad g_2^3 - 27g_3^2 \neq 0.$$

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,

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

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  $\tau$  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*,

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

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. 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*

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

The product is taken over primes  $p$ , and  $1_N(p)$  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, analogous to a complex elliptic curve in that both are complex tori and thus have 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,

$$J_0(N) \longrightarrow E.$$

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  $g_2, g_3 \in \mathbf{Q}$ . Modular curves, Jacobians, and Abelian varieties are similarly associated to solution sets of systems of polynomial equations over  $\mathbf{Q}$ , 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  $\mathbf{Q}$ ,

$$X_0(N)_{\text{alg}} \longrightarrow E, \quad J_0(N)_{\text{alg}} \longrightarrow E, \quad A_{f,\text{alg}} \longrightarrow E.$$

We discuss only the first of these in detail since  $X_0(N)_{\text{alg}}$  is a curve while  $J_0(N)_{\text{alg}}$  and  $A_{f,\text{alg}}$  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 a positive integer  $N$  called the *conductor* of  $E$ ,



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

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  $\mathbf{C}$ . 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  $\ell$ -adic Galois representations, certain homomorphisms of Galois groups into matrix groups. Such representations are associated to elliptic curves and to modular forms, incorporating the ideas from Chapters 6 through 8 into a framework with rich additional algebraic structure. The corresponding version of the Modularity Theorem is: *Every Galois representation associated to an elliptic curve over  $\mathbf{Q}$  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 two broader conjectures that Galois representations arise from 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. Comments and corrections should be sent to the second author at [jerry@reed.edu](mailto:jerry@reed.edu).

*July 2004*

Fred Diamond  
*Brandeis University*  
*Waltham, MA*

Jerry Shurman  
*Reed College*  
*Portland, OR*

---

# Contents

<b>Preface</b> .....	vii
<b>1 Modular Forms, Elliptic Curves, and Modular Curves</b> .....	1
1.1 First definitions and examples .....	1
1.2 Congruence subgroups .....	11
1.3 Complex tori .....	24
1.4 Complex tori as elliptic curves .....	31
1.5 Modular curves and moduli spaces .....	37
<b>2 Modular Curves as Riemann Surfaces</b> .....	45
2.1 Topology .....	45
2.2 Charts .....	48
2.3 Elliptic points .....	52
2.4 Cusps .....	57
2.5 Modular curves and Modularity .....	63
<b>3 Dimension Formulas</b> .....	65
3.1 The genus .....	65
3.2 Automorphic forms .....	71
3.3 Meromorphic differentials .....	77
3.4 Divisors and the Riemann–Roch Theorem .....	83
3.5 Dimension formulas for even $k$ .....	85
3.6 Dimension formulas for odd $k$ .....	89
3.7 More on elliptic points .....	92
3.8 More on cusps .....	98
3.9 More dimension formulas .....	106
<b>4 Eisenstein Series</b> .....	109
4.1 Eisenstein series for $SL_2(\mathbf{Z})$ .....	109
4.2 Eisenstein series for $\Gamma(N)$ when $k \geq 3$ .....	111
4.3 Dirichlet characters, Gauss sums, and eigenspaces .....	116

4.4	Gamma, zeta, and $L$ -functions .....	120
4.5	Eisenstein series for the eigenspaces when $k \geq 3$ .....	126
4.6	Eisenstein series of weight 2 .....	130
4.7	Bernoulli numbers and the Hurwitz zeta function .....	133
4.8	Eisenstein series of weight 1 .....	138
4.9	The Fourier transform and the Mellin transform .....	143
4.10	Nonholomorphic Eisenstein series .....	147
4.11	Modular forms via theta functions .....	155
<b>5</b>	<b>Hecke Operators</b> .....	<b>163</b>
5.1	The double coset operator .....	163
5.2	The $\langle d \rangle$ and $T_p$ operators .....	168
5.3	The $\langle n \rangle$ and $T_n$ operators .....	178
5.4	The Petersson inner product .....	181
5.5	Adjoints of the Hecke Operators .....	183
5.6	Oldforms and Newforms .....	187
5.7	The Main Lemma .....	189
5.8	Eigenforms .....	195
5.9	The connection with $L$ -functions .....	200
5.10	Functional equations .....	204
5.11	Eisenstein series again .....	205
<b>6</b>	<b>Jacobians and Abelian Varieties</b> .....	<b>211</b>
6.1	Integration, homology, the Jacobian, and Modularity .....	212
6.2	Maps between Jacobians .....	217
6.3	Modular Jacobians and Hecke operators .....	226
6.4	Algebraic numbers and algebraic integers .....	230
6.5	Algebraic eigenvalues .....	233
6.6	Eigenforms, Abelian varieties, and Modularity .....	240
<b>7</b>	<b>Modular Curves as Algebraic Curves</b> .....	<b>249</b>
7.1	Elliptic curves as algebraic curves .....	250
7.2	Algebraic curves and their function fields .....	257
7.3	Divisors on curves .....	268
7.4	The Weil pairing algebraically .....	275
7.5	Function fields over $\mathbf{C}$ .....	279
7.6	Function fields over $\mathbf{Q}$ .....	287
7.7	Modular curves as algebraic curves and Modularity .....	290
7.8	Isogenies algebraically .....	295
7.9	Hecke operators algebraically .....	300
<b>8</b>	<b>The Eichler–Shimura Relation and <math>L</math>-functions</b> .....	<b>309</b>
8.1	Elliptic curves in arbitrary characteristic .....	310
8.2	Algebraic curves in arbitrary characteristic .....	317
8.3	Elliptic curves over $\mathbf{Q}$ and their reductions .....	322

8.4	Elliptic curves over $\overline{\mathbf{Q}}$ and their reductions	329
8.5	Reduction of algebraic curves and maps	336
8.6	Modular curves in characteristic $p$ : Igusa's Theorem	347
8.7	The Eichler–Shimura Relation	349
8.8	Fourier coefficients, $L$ -functions, and Modularity	356
<b>9</b>	<b>Galois Representations</b>	<b>365</b>
9.1	Galois number fields	366
9.2	The $\ell$ -adic integers and the $\ell$ -adic numbers	372
9.3	Galois representations	376
9.4	Galois representations and elliptic curves	382
9.5	Galois representations and modular forms	386
9.6	Galois representations and Modularity	391
	<b>Hints and Answers to the Exercises</b>	<b>401</b>
	<b>List of Symbols</b>	<b>421</b>
	<b>Index</b>	<b>427</b>
	<b>References</b>	<b>433</b>

# Modular Forms, Elliptic Curves, and Modular Curves

This chapter introduces three central objects of the book.

*Modular forms* are functions on the complex upper half plane. A matrix group called the modular group acts on the upper half plane, and modular forms are the functions that transform in a nearly invariant way under the action and satisfy a holomorphy condition. Restricting the action to subgroups of the modular group called congruence subgroups gives rise to more modular forms.

A *complex elliptic curve* is a quotient of the complex plane by a lattice. As such it is an Abelian group, a compact Riemann surface, a torus, and—nonobviously—in bijective correspondence with the set of ordered pairs of complex numbers satisfying a cubic equation of the form  $E$  in the preface.

A *modular curve* is a quotient of the upper half plane by the action of a congruence subgroup. That is, two points are considered the same if the group takes one to the other.

These three kinds of object are closely related. Modular curves are mapped to by *moduli spaces*, equivalence classes of complex elliptic curves enhanced by associated torsion data. Thus the points of modular curves represent enhanced elliptic curves. Consequently, functions on the moduli spaces satisfying a homogeneity condition are essentially the same thing as modular forms.

Related reading: Gunning [Gun62], Koblitz [Kob93], Schoeneberg [Sch74], and Chapter 7 of Serre [Ser73] are standard first texts on this subject. For modern expositions of classical modular forms in action see [Cox84] (reprinted in [BBB00]) and [Cox97].

## 1.1 First definitions and examples

The *modular group* is the group of 2-by-2 matrices with integer entries and determinant 1,

$$\mathrm{SL}_2(\mathbf{Z}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a, b, c, d \in \mathbf{Z}, ad - bc = 1 \right\}.$$

The modular group is generated by the two matrices

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

(Exercise 1.1.1). Each element of the modular group is also viewed as an automorphism (invertible self-map) of the Riemann sphere  $\widehat{\mathbf{C}} = \mathbf{C} \cup \{\infty\}$ , the fractional linear transformation

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}(\tau) = \frac{a\tau + b}{c\tau + d}, \quad \tau \in \widehat{\mathbf{C}}.$$

This is understood to mean that if  $c \neq 0$  then  $-d/c$  maps to  $\infty$  and  $\infty$  maps to  $a/c$ , and if  $c = 0$  then  $\infty$  maps to  $\infty$ . The identity matrix  $I$  and its negative  $-I$  both give the identity transformation, and more generally each pair  $\pm\gamma$  of matrices in  $\mathrm{SL}_2(\mathbf{Z})$  gives a single transformation. The group of transformations defined by the modular group is generated by the maps described by the two matrix generators,

$$\tau \mapsto \tau + 1 \quad \text{and} \quad \tau \mapsto -1/\tau.$$

The *upper half plane* is

$$\mathcal{H} = \{\tau \in \mathbf{C} : \mathrm{Im}(\tau) > 0\}.$$

Readers with some background in Riemann surface theory—which is not necessary to read this book—may recognize  $\mathcal{H}$  as one of the three simply connected Riemann surfaces, the other two being the plane  $\mathbf{C}$  and the sphere  $\widehat{\mathbf{C}}$ . The formula

$$\mathrm{Im}(\gamma(\tau)) = \frac{\mathrm{Im}(\tau)}{|c\tau + d|^2}, \quad \gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbf{Z})$$

(Exercise 1.1.2(a)) shows that if  $\gamma \in \mathrm{SL}_2(\mathbf{Z})$  and  $\tau \in \mathcal{H}$  then also  $\gamma(\tau) \in \mathcal{H}$ , i.e., the modular group maps the upper half plane back to itself. In fact the modular group acts on the upper half plane, meaning that  $I(\tau) = \tau$  where  $I$  is the identity matrix (as was already noted) and  $(\gamma\gamma')(\tau) = \gamma(\gamma'(\tau))$  for all  $\gamma, \gamma' \in \mathrm{SL}_2(\mathbf{Z})$  and  $\tau \in \mathcal{H}$ . This last formula is easy to check (Exercise 1.1.2(b)).

**Definition 1.1.1.** Let  $k$  be an integer. A meromorphic function  $f : \mathcal{H} \rightarrow \mathbf{C}$  is *weakly modular of weight  $k$*  if

$$f(\gamma(\tau)) = (c\tau + d)^k f(\tau) \quad \text{for } \gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \text{ and } \tau \in \mathcal{H}.$$

Section 1.2 will show that if this transformation law holds when  $\gamma$  is each of the generators  $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  and  $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$  then it holds for all  $\gamma \in \mathrm{SL}_2(\mathbf{Z})$ . In other words,  $f$  is weakly modular of weight  $k$  if

$$f(\tau + 1) = f(\tau) \quad \text{and} \quad f(-1/\tau) = \tau^k f(\tau).$$

Weak modularity of weight 0 is simply  $\mathrm{SL}_2(\mathbf{Z})$ -invariance,  $f \circ \gamma = f$  for all  $\gamma \in \mathrm{SL}_2(\mathbf{Z})$ . Weak modularity of weight 2 is also natural: complex analysis relies on path integrals of differentials  $f(\tau)d\tau$ , and  $\mathrm{SL}_2(\mathbf{Z})$ -invariant path integration on the upper half plane requires such differentials to be invariant when  $\tau$  is replaced by any  $\gamma(\tau)$ . But (Exercise 1.1.2(c))

$$d\gamma(\tau) = (c\tau + d)^{-2} d\tau,$$

and so the relation  $f(\gamma(\tau))d(\gamma(\tau)) = f(\tau)d\tau$  is

$$f(\gamma(\tau)) = (c\tau + d)^2 f(\tau),$$

giving Definition 1.1.1 with weight  $k = 2$ . Weight 2 will play an especially important role later in this book since it is the weight of the modular form in the Modularity Theorem. The weight 2 case also leads inexorably to higher even weights—multiplying two weakly modular functions of weight 2 gives a weakly modular function of weight 4, and so on. Letting  $\gamma = -I$  in Definition 1.1.1 gives  $f = (-1)^k f$ , showing that the only weakly modular function of any odd weight  $k$  is the zero function, but nonzero odd weight examples exist in more general contexts to be developed soon. Another motivating idea for weak modularity is that while it does not make a function  $f$  fully  $\mathrm{SL}_2(\mathbf{Z})$ -invariant, at least  $f(\tau)$  and  $f(\gamma(\tau))$  always have the same zeros and poles since the factor  $c\tau + d$  on  $\mathcal{H}$  has neither.

Modular forms are weakly modular functions that are also holomorphic on the upper half plane and holomorphic at  $\infty$ . To define this last notion, recall that  $\mathrm{SL}_2(\mathbf{Z})$  contains the translation matrix

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} : \tau \mapsto \tau + 1,$$

for which the factor  $c\tau + d$  is simply 1, so that  $f(\tau + 1) = f(\tau)$  for every weakly modular function  $f : \mathcal{H} \rightarrow \mathbf{C}$ . That is, weakly modular functions are  $\mathbf{Z}$ -periodic. Let  $D = \{q \in \mathbf{C} : |q| < 1\}$  be the open complex unit disk, let  $D' = D - \{0\}$ , and recall from complex analysis that the  $\mathbf{Z}$ -periodic holomorphic map  $\tau \mapsto e^{2\pi i \tau} = q$  takes  $\mathcal{H}$  to  $D'$ . Thus, corresponding to  $f$ , the function  $g : D' \rightarrow \mathbf{C}$  where  $g(q) = f(\log(q)/(2\pi i))$  is well defined even though the logarithm is only determined up to  $2\pi i \mathbf{Z}$ , and  $f(\tau) = g(e^{2\pi i \tau})$ . If  $f$  is holomorphic on the upper half plane then the composition  $g$  is holomorphic on the punctured disk since the logarithm can be defined holomorphically about each point, and so  $g$  has a Laurent expansion  $g(q) = \sum_{n \in \mathbf{Z}} a_n q^n$  for  $q \in D'$ . The relation  $|q| = e^{-2\pi \mathrm{Im}(\tau)}$  shows that  $q \rightarrow 0$  as  $\mathrm{Im}(\tau) \rightarrow \infty$ . So, thinking of  $\infty$  as lying far in the imaginary direction, define  $f$  to be *holomorphic at  $\infty$*  if  $g$  extends holomorphically to the puncture point  $q = 0$ , i.e., the Laurent series sums over  $n \in \mathbf{N}$ . This means that  $f$  has a *Fourier expansion*



$$f(\tau) = \sum_{n=0}^{\infty} a_n(f) q^n, \quad q = e^{2\pi i \tau}.$$

Since  $q \rightarrow 0$  if and only if  $\text{Im}(\tau) \rightarrow \infty$ , showing that a weakly modular holomorphic function  $f: \mathcal{H} \rightarrow \mathbb{C}$  is holomorphic at  $\infty$  doesn't require computing its Fourier expansion, only showing that  $\lim_{\text{Im}(\tau) \rightarrow \infty} f(\tau)$  exists or even just that  $f(\tau)$  is bounded as  $\text{Im}(\tau) \rightarrow \infty$ .

**Definition 1.1.2.** Let  $k$  be an integer. A function  $f: \mathcal{H} \rightarrow \mathbb{C}$  is a **modular form of weight  $k$**  if

- (1)  $f$  is holomorphic on  $\mathcal{H}$ ,
- (2)  $f$  is weakly modular of weight  $k$ ,
- (3)  $f$  is holomorphic at  $\infty$ .

The set of modular forms of weight  $k$  is denoted  $\mathcal{M}_k(\text{SL}_2(\mathbb{Z}))$ .

It is easy to check that  $\mathcal{M}_k(\text{SL}_2(\mathbb{Z}))$  forms a vector space over  $\mathbb{C}$  (Exercise 1.1.3(a)). Holomorphy at  $\infty$  will make the dimension of this space, and of more spaces of modular forms to be defined in the next section, finite. We will compute many dimension formulas in Chapter 3. When  $f$  is holomorphic at  $\infty$  it is tempting to define  $f(\infty) = g(0) = a_0$ , but the next section will show that this doesn't work in a more general context.

The product of a modular form of weight  $k$  with a modular form of weight  $l$  is a modular form of weight  $k+l$  (Exercise 1.1.3(b)). Thus the sum

$$\mathcal{M}(\text{SL}_2(\mathbb{Z})) = \bigoplus_{k \in \mathbb{Z}} \mathcal{M}_k(\text{SL}_2(\mathbb{Z}))$$

forms a ring, a so-called graded ring because of its structure as a sum.

The zero function on  $\mathcal{H}$  is a modular form of every weight, and every constant function on  $\mathcal{H}$  is a modular form of weight 0. For nontrivial examples of modular forms, let  $k > 2$  be an even integer and define the *Eisenstein series of weight  $k$*  to be a 2-dimensional analog of the Riemann zeta function  $\zeta(k) = \sum_{d=1}^{\infty} 1/d^k$ ,

$$G_k(\tau) = \sum'_{(c,d)} \frac{1}{(c\tau + d)^k}, \quad \tau \in \mathcal{H},$$

where the primed summation sign means to sum over nonzero integer pairs  $(c, d) \in \mathbb{Z}^2 - \{(0, 0)\}$ . The sum is absolutely convergent and converges uniformly on compact subsets of  $\mathcal{H}$  (Exercise 1.1.4(c)), so  $G_k$  is holomorphic on  $\mathcal{H}$  and its terms may be rearranged. For any  $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \text{SL}_2(\mathbb{Z})$ , compute that