Ze-Li Dou
Qiao Zhang

Six Short Chapters on Automorphic Forms and *L*-functions

（自守形式与*L*-函数简明六章）

SCIENCE PRESS
Beijing

Six Short Chapters on Automorphic Forms and *L*-functions

(自守形式与 *L*-函数简明六章)

Ze-Li Dou and Qiao Zhang

SCIENCE PRESS
Beijing

Responsible Editors: Liping Wang and Yang Fang

Published by Science Press
16 Donghuangchenggen North Street
Beijing 100717, P. R. China

Printed in Beijing

ISBN 978-7-03-033168-7

To our parents,

Fukang Dou, Yuzhen Pan Dou,

Yegang Zhang, Cuiying Sun

谁言寸草心
报得三春晖*

But how could thanks from verdure's meanest blade
Indulgent sun and gentle spring repay?

* Meng Jiao (751-814, Tang Dynasty), *You Zi Yin* (in Chinese)

Preface

This modest tract offers a little light reading on a heavy subject. In six short and relatively independent chapters, we provide motivated discussion on a few topics in number theory that are tremendously active and rapidly progressing today.

Both the content and style of this book had their origin in invited talks we gave in the last few years, in China, Turkey, and the United States. Our audiences consisted in mathematicians and graduate students whose backgrounds were rather diverse. Although all were interested in the topics in general, many were not experts in the field. It was natural, therefore, that we endeavored to give motivation on each main topic before delving into deeper material. Also, we attempted to keep the lectures as independent from one another as possible, and to make at least part of each lecture comprehensible to the entire audience. Finally, in order not to have the main ideas eclipsed by the subordinate (but often heavily technical) material, as a rule we omitted detailed proofs. Of course, we should point out that the lectures do form an integral whole as well: they evolve around the moment conjecture for L-functions and the period conjectures of Shimura.

When the invitation to provide a written account of these lectures came, we at first felt the temptation of generating a fuller and more coherent work. Upon further consideration, however, we have decided against this idea. To undertake such a project would require a vast amount of time and energy, and we are not certain that we can afford to do so at the present time. On the other hand, we also believe that a motivated—and as nontechnical as possible—introduction of the main topics can be useful, since it can help bridge the gap between the basics and certain specialized research areas, and perhaps also to serve as a road map for relevant literature. Consequently, what the reader now has in hand is an extended version of the original talks: revised, somewhat fleshed out, but otherwise retaining most of the characteristics of the original lectures themselves. Although this work is by no means complete or comprehensive, it is our hope that, through this, the reader can gain an overall appreciation for the main topics and conjectures considered herein. In the meantime, to the reader who intends to study these topics in depth, we hope that we have offered a motivated guide to the literature.

A rough idea for the content and organization of this book can easily be gained by glancing at the table of content, and we shall therefore not elaborate too much here. During the preparation for our lectures as well as for this manuscript, we have benefited from many existing resources in the literature. For the more introductory parts (which form the early parts of most of the chapters), we wish to mention especially Goro Shimura's *Introduction to the Arithmetic Theory of Automorphic Functions*, Danie L Bump's *Automorphic Forms and Representations*, and Henryk Iwaniec and Emmanuel Kowalski's *Analytic Number Theory*. In several places, the arrangement of concepts and the choice of certain examples have been influenced by these excellent books. It is impossible to acknowledge in detail the numerous works we have consulted for the rest of the book, and the reader must be referred to the Bibliography. Some work by the authors themselves on the conjectures have been briefly (and incompletely) summarized here as well, mainly in the two final chapters.

It remains our most pleasant duty to acknowledge the many people from whom we have received much help and support, mathematically and otherwise. First and foremost, we thank Professor Goro Shimura and Professor Dorian Goldfeld, our thesis advisors. We are grateful to Professor Shou-Wu Zhang of Columbia University, Professor Hongwen Lu of Tongji University, Professor K. Ilhan Ikeda of Istanbul Bilgi University, Professor Juping Wang of Fudan University, and Professor Tianze Wang of Henan University for their kind invitations and their steadfast support. We are much indebted to Professor Paula Cohen Tretkoff of Texas A&M University and Professor Yingchun Cai of Tongji University for their encouragement and advice. With gratitude we acknowledge the kindly and efficient assistance from the editors, Liping Wang and Lisa Libin Fan. Finally, we give our heartfelt thanks to all members of our families, and especially to our wives, Susan Staples and Jing Tian, for their longsuffering and their love.

Ze-Li Dou and Qiao Zhang
Fort Worth, TX, USA
October 11, 2011

Contents

Chapter 1

Modular forms and the Shimura-Taniyama Conjecture

The concept of modular form are based on very natural considerations. In this chapter we recount some rudiments of the theory of modular forms without assuming any previous knowledge of the subject on the reader's part. The number theoretic interest of the subject becomes apparent when we describe the Hecke operators on the spaces of modular forms and the L-functions attached to eigenforms. The connection between elliptic curves and modular forms of weight 2 is briefly described towards the end in order to state the celebrated Shimura–Taniyama Conjecture, which is now a theorem of A. Wiles, et al. See [Wi95] and related articles.

The standard reference for the foundational material of this subject is the book *Introduction to the Arithmetic Theory of Automorphic Functions* by G. Shimura [Sh71]. Other excellent textbooks and references are also readily available; as a somewhat random sampling we mention [Bo97], [Bu97], [Ge75], [Gu62], [Iw02], [Kn92], [Mi89], and [Se73].

1.1 Elliptic functions

We begin with a lattice on the complex plane, $\mathbb{C}$. Let ω_1 and ω_2 be two nonzero complex numbers that are linearly independent over $\mathbb{R}$. In other words, we suppose that the points 0, ω_1, and ω_2 do not lie on a straight line. Then the set of all linear combinations of ω_1 and ω_2 with integer coefficients is a *lattice*

$$L = L(\omega_1, \omega_2) = \{m\omega_1 + n\omega_2 : m, n \in \mathbb{Z}\}. \tag{1.1}$$

We seek to construct functions which are doubly periodic with respect to L, that is, functions f such that

$$f(z+\omega) = f(z), \quad \forall \omega \in L. \tag{1.2}$$

According to Liouville's Theorem, every bounded entire function must be constant, hence non-constant doubly periodic functions cannot be holomorphic everywhere. We therefore seek meromorphic functions with that property.

Definition 1.1.1 *An elliptic function with respect to a lattice L is a meromorphic function f such that* (1.2) *holds.*

The simplest non-constant elliptic function is the *Weierstrass ℘-function*, defined by

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in L-\{0\}} \left(\frac{1}{(z+\omega)^2} - \frac{1}{\omega^2} \right). \tag{1.3}$$

This function is absolutely convergent except when $z \in L$, where it has a double pole.

The derivative of $\wp$ is another elliptic function:

$$\wp'(z) = -2 \sum_{\omega \in L} \frac{1}{(z+\omega)^3}. \tag{1.4}$$

This time we have triple poles at lattice points.

Though linearly independent, the Weierstrass functions $\wp$ and $\wp'$ are algebraically related. We can see this by comparing their Laurent series. Let us introduce a notation, $\mathcal{G}_{2k}$:

$$\mathcal{G}_{2k}(L) = \sum_{\omega \in L-\{0\}} \frac{1}{\omega^{2k}}, \quad \forall 2 \leqslant k \in \mathbb{Z}. \tag{1.5}$$

Then it can be checked that we have

$$\wp(z) = \frac{1}{z^2} + \sum_{k=1}^{\infty} 2k(2k+1)\mathcal{G}_{2k+2}(L) z^{2k-1} \tag{1.6}$$

and

$$\wp'(z) = -\frac{2}{z^3} + \sum_{k=1}^{\infty} (2k+1)\mathcal{G}_{2k+2}(L) z^{2k}. \tag{1.7}$$

It follows from these equations that $\wp$ and $\wp'$ satisfy the relation

$$(\wp')^2 = 4\wp^3 - 60\mathcal{G}_4(L)\wp - 140\mathcal{G}_6(L). \tag{1.8}$$

To see that this identity holds, one needs only write out the respective terms in (1.8) for the first few terms, and check that

$$(\wp')^2 - 4\wp^3 + 60\mathcal{G}_4(L)\wp = 140\mathcal{G}_6(L) + \text{ higher-order terms.}$$

But the right-hand side no longer has a pole, while the left-hand side is clearly doubly periodic. From this we deduce that the assertion is true.

It is standard to write

$$g_2 = 60\mathcal{G}_4 \quad \text{and} \quad g_3 = 140\mathcal{G}_6. \tag{1.9}$$

Then (1.8) assumes the simpler form

$$(\wp')^2 = 4\wp^3 - g_2\wp - g_3. \tag{1.10}$$

Given L, the elliptic functions with respect to L form a field. Because of the relation (1.8), one can prove by using the Riemann-Roch Theorem that this field is precisely $\mathbb{C}(\wp, \wp')$.

1.2 Modular forms

If a lattice L is fixed, then the relation between $\wp$ and $\wp'$ is precisely known, once $\mathcal{G}_4(L)$ and $\mathcal{G}_6(L)$ have been evaluated. We may naturally ask how the relation changes when one varies L. We pose the question more generally as follows.

Question Given $k \geqslant 2$, what is the behavior of $\mathcal{G}_{2k}$ as a function of the lattice L?

It is obvious from the definition of $\mathcal{G}_{2k}$ that we have

$$\mathcal{G}_{2k}(zL) = z^{-2k}\mathcal{G}_{2k}(L). \tag{1.11}$$

Thus there is some rigidity regarding the re-scaling of the lattice. This makes it possible for us to interpret the behavior of $\mathcal{G}_{2k}(L)$ via a function of one complex variable. Note that

$$L(\omega_1, \omega_2) = \omega_2 L\Big(\frac{\omega_1}{\omega_2}, 1\Big). \tag{1.12}$$

Without loss of generality, we may assume that

$$\frac{\omega_1}{\omega_2} \in H \stackrel{\text{def}}{=} \text{the upper-half plane} = \{z \in \mathbb{C} : \text{Im}(z) > 0\}. \tag{1.13}$$

Then we define

$$G_{2k}(z) = \mathcal{G}_{2k}(L(z,1)). \tag{1.14}$$

In general, therefore, if $L = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$, and $z = \omega_1/\omega_2 \in H$, then

$$\mathcal{G}_{2k}(L) = \omega_2^{-2k} G_{2k}(z). \tag{1.15}$$

There are further invariance properties. The function $\mathcal{G}_{2k}$, being a function of the lattice L, is clearly invariant under different choices of bases. However, when the generators of L, ω_1 and ω_2, are chosen differently, the quotient $z = \omega_1/\omega_2$ does vary. For example, the vectors $\omega_1 + \omega_2$ and ω_2 also generate L. Thus we must have $G_{2k}(z) = G_{2k}(z+1)$.

In general, suppose

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}). \tag{1.16}$$

Then the lattice $L(z,1)$, generated by $z \in H$ and 1, is also generated by

$$\alpha \begin{pmatrix} z \\ 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z \\ 1 \end{pmatrix} = \begin{pmatrix} az+b \\ cz+d \end{pmatrix}.$$

Thus

$$L(z,1) = L(az+b, cz+d). \tag{1.17}$$

Recall the standard notation for linear fractional transformations: if $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then

$$\alpha z = \frac{az+b}{cz+d}.$$

It is easy to check that, if $\alpha \in \mathrm{SL}_2(\mathbb{Z})$, then α maps H onto itself.

Let us rewrite (1.17) in terms of G_{2k}. By (1.15), we have

$$\mathcal{G}_{2k}(L(z,1)) = G_{2k}(z), \quad \mathcal{G}_{2k}(L(az+b, cz+d)) = (cz+d)^{-2k} G_{2k}(\alpha z).$$

Therefore

$$G_{2k}(\alpha z) = (cz+d)^{2k} G_{2k}(z). \tag{1.18}$$

This is the main defining property for modular forms.

Definition 1.2.1 *A modular form of (even) weight k with respect to $\mathrm{SL}_2(\mathbb{Z})$ is a holomorphic function $f : H \to \mathbb{C}$ such that*

(1) $f(\alpha z) = (cz+d)^k f(z), \quad \forall \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z});$

(2) *f is "holomorphic at ∞," that is, f has a Fourier expansion of the form*

$$f(z) = \sum_{n=0}^{\infty} a_n e^{2\pi i n z}.$$

(*In other words, there are no negative indices.*)

We call $\mathrm{SL}_2(\mathbb{Z})$ the *full modular group.* Its subgroups of finite index are called *modular groups.*

Note that

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \quad \text{and} \quad \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

The first matrix yields the identity $f(z+1) = f(z)$, which has already been observed, and guarantees a Fourier expansion for the modular form. Thus condition (2) of the definition can be stated in the way we have done. The fact that the second matrix belongs to $\mathrm{SL}_2(\mathbb{Z})$ implies that a modular form for the full modular group cannot have odd weight, unless it is 0. For this reason only even weights have been mentioned in the definition.

It is convenient to introduce the notation

$$q = e^{2\pi i z}.$$

With the q notation, the Fourier expansion of a modular form f acquires a neater looking form

$$f(z) = \sum_{n=0}^{\infty} a_n q^n.$$

1.3 Examples

It is not surprising that the functions $G_{2k}(z)$ for $k \geqslant 2$ all turn out to be modular forms. They are called *Eisenstein series.* The weight of $G_{2k}(z)$ is $2k$. In fact the following identity is true:

$$G_{2k}(z) = 2\zeta(k) + \frac{2(2\pi i)^{2k}}{(2k-1)} \sum_{n=1}^{\infty} \sigma_{2k-1}(n) q^n. \tag{1.19}$$

Here ζ is the *Riemann zeta function*, and the symbol σ_m is defined by

$$\sigma_m(n) = \sum_{d|n} d^m.$$

In general, if f and g are two modular forms of weight k and $a, b \in \mathbb{C}$ are arbitrary constants, then the linear combination $af + bg$ is another modular form of weight k. It follows that the set of all modular forms of a given weight k is a vector space. We denote this space by $\mathcal{M}_k$.

It is also easy to verify that, if f and g are modular forms of weights k and ℓ, respectively, then their product fg is a modular form of weight $k+\ell$. (Heuristically, therefore, we may expect f/g to be a modular form of weight $k - \ell$ if $k > \ell$. This is of course not precisely true, but it provides a helpful perspective nevertheless.)

Definition 1.3.1 *A modular form f that "vanishes at ∞," that is, a form with Fourier expansion*

$$f(z) = \sum_{n=1}^{\infty} a_n q^n$$

is called a cusp form.

Clearly, linear combinations of cusp forms are again cusp forms. Thus the set of all weight k cusp forms also form a vector space. It is denoted by $\mathcal{S}_k$.

From (1.19) we see that none of the Eisenstein series is a cusp form. However, it is quite easy to manufacture one, since we could simply take two modular forms (say Eisenstein series), match up their weights, and then cancel out the leading coefficient.

More concretely, we define

$$\Delta = g_2^3 - 27g_3^2. \tag{1.20}$$

Since g_2 has weight 4 and g_3 has weight 6, Δ is a modular form of weight 12. A closer look discovers the fact that the constant term for Δ is 0. Therefore Δ is a cusp form.

Definition 1.3.2 *A meromorphic function $f : H \to \mathbb{C}$ such that*

$$f(\alpha z) = f(z), \quad \forall \alpha \in \mathrm{SL}_2(\mathbb{Z})$$

is called a modular function.

Thus a modular function is a "modular form of weight 0". However, we have to sacrifice holomorphicity.

The cusp form Δ never assumes the value 0. This allows us to define

$$j = \frac{1728 g_2^3}{\Delta}.$$

This *j-function* is a modular function. (Heuristically, $12 - 12 = 0$.) More precisely, one can show that $j(z)$ has a Fourier expansion with integer coefficients. Therefore, if we view j as a function defined on the quotient space $\mathrm{SL}_2(\mathbb{Z})\backslash H^*$, where $H^* = H \cup \{\infty\}$, then it has just one simple pole, located at ∞.

We note that, in general, the definiton of H^* depends on the group in question. That is, given a Fuchsian group Γ, H^* is the union of H with the set of cusps of Γ. This will be relevant later.

Although the examples we have given so far suffice to illustrate the definitions of modular forms, cusp forms and modular functions, we may legitimately ask whether or not they also serve as typical or representative examples of these concepts. The surprising answer is that they are quite adequate for the construction of all modular forms and modular functions on the full modular group.

Since $H^* = H \cup \{\infty\}$ has genus 0 and j has just a simple pole, the Riemann-Roch Theorem implies that the field of modular functions is $\mathbb{C}(j)$. Thus j suffices for the characterization of modular functions.

The Riemann-Roch Theorem also allows us to conclude that $\dim(\mathcal{M}_2) = 0$, $\dim(\mathcal{S}_k) = 0$ for $k = 2, 4, 6, 8$, and 10, and

$$\mathcal{S}_k = \Delta \cdot \mathcal{M}_{k-12}. \tag{1.21}$$

On the other hand, given $f \in \mathcal{M}_k$ for an even $k \geqslant 4$, there clearly exists a constant $c \in \mathbb{C}$ such that $f - cG_k \in \mathcal{S}_k$. So

$$\mathcal{M}_k = \mathbb{C}G_k \oplus \mathcal{S}_k, \quad \forall \text{ even } k \geqslant 4. \tag{1.22}$$

These facts suffice to describe all modular forms with respect to the full modular group $\mathrm{SL}_2(\mathbb{Z})$. To illustrate this with a simple example, let us compute the dimension of $\mathcal{M}_{28}$. By (1.22), we have $\mathcal{M}_{28} = \mathbb{C} \cdot G_{28} \oplus \mathcal{S}_{28}$. But $\mathcal{S}_{28} = \Delta \cdot \mathcal{M}_{16}$ by (1.21). This shows that $\dim(\mathcal{M}_{28}) = 1 + \dim(\mathcal{M}_{16})$. A similar analysis of $\mathcal{M}_{16}$ leads to the conclusion $\dim(\mathcal{M}_{16}) = 1 + \dim(\mathcal{M}_4)$. But $\mathcal{M}_4 = \mathbb{C} \cdot G_4$, which is of dimension 1. Thus $\dim(\mathcal{M}_{28}) = 1 + 1 + 1 = 3$.

By induction, the above discussion shows that every element of $\mathcal{M}_k$ is generated algebraically by the three forms G_4, G_6, and Δ. In fact, by (1.20), even Δ can be dispensed with! General formulas for the dimensions of $\mathcal{M}_k$ and $\mathcal{S}_k$ can be easily derived from (1.21) and (1.22), and can be found in almost all textbooks. We therefore omit them in the interest of space.

1.4 Hecke operators and eigenforms

Thus far we have treated modular forms as analytic objects. As such they arise from very natural considerations. Their number theoretic interest, however, becomes immediately apparent when we consider the Hecke operators. We shall restrict ourselves to cusp forms in this section.

The Hecke operators are linear operators on the vector spaces $\mathcal{S}_k$, which can be made into a Hilbert space by the introduction of an inner product, called the *Petersson inner product*. Given $f, g \in \mathcal{S}_k$, it is defined by

$$\langle f, g \rangle = \int_{\mathrm{SL}_2(\mathbb{Z})\backslash H} f(z)\overline{g(z)} y^k \frac{\mathrm{d}x\mathrm{d}y}{y^2}.$$

We omit the verification of convergence here.

There is a *Hecke operator* for every positive integer $1 \leqslant m \in \mathbb{Z}$; it is denoted by $T(m)$. We take the utilitarian approach of immediately exhibiting some of their basic properties.

(1) The Hecke operators are commutative. That is,

$$T(m)T(n) = T(n)T(m), \quad \forall 1 \leqslant m, n \in \mathbb{Z}. \tag{1.23}$$

Here $T(m)T(n)$ means the composition of the two operators.

(2) The Hecke operators are multiplicative. More precisely, we have the following two formulas:

$$T(m)T(n) = T(mn), \quad \text{if } (m, n) = 1, \text{ and} \tag{1.24}$$

$$T(p^{r+1}) = T(p)T(p^r) - p^{k-1}T(p^{r-1}), \quad \text{if } p \text{ is prime.} \tag{1.25}$$

(3) The Hecke operators are self-adjoint with respect to the Petersson inner product. Thus

$$\langle f|T(m), g \rangle = \langle f, g|T(m) \rangle, \tag{1.26}$$

where $f|T(m)$ denotes the image of f under $T(m)$.