1
Introduction

Integers form a group under addition but not under multiplication. However, matrices with integer entries can form groups under multiplication. For example, 2×2 matrices of unit determinant and integer entries $a, b, c, d \in \mathbb{Z}$,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad\qquad ad - bc = 1$$

form the group $\mathrm{SL}(2, \mathbb{Z})$ under multiplication, referred to as the *modular group*. Functions and differential forms that are invariant under $\mathrm{SL}(2, \mathbb{Z})$ are referred to as *modular functions* and *modular forms*, respectively. Modular functions generalize *periodic functions* and *elliptic functions*, to which they are intimately related. In turn, modular functions and modular forms are special cases of *automorphic functions* and *automorphic forms* which are invariant under more general infinite discrete subgroups of $\mathrm{SL}(2, \mathbb{R})$, or under more general multiplicative groups of matrices with integer entries such as $\mathrm{SL}(m, \mathbb{Z})$ and $\mathrm{Sp}(2m, \mathbb{Z})$, referred to as *arithmetic groups*.

In mathematics, the theory of elliptic functions was developed by Gauss, Jacobi, and Weierstrass, building on the study of elliptic integrals by Euler, Legendre, and Abel, and was motivated in part by questions ranging from number theory to the solvability of algebraic equations by radicals. Riemann generalized elliptic functions to Riemann surfaces of arbitrary genus.

The development of modular functions and forms dates back to Eisenstein, Kronecker, and Hecke. Automorphic functions and forms were studied by Fuchs, Fricke, Klein, and Poincaré. In modern times, among many other developments, the Taniyama–Shimura–Weil conjecture ultimately led to the proof of Fermat's Last Theorem by Wiles and Taylor and to a proof of the Modularity Theorem by Breuil, Conrad, Diamond, and Taylor.

A fundamental role was played by modular forms and quasi-modular forms in the solution to the sphere packing problem in eight dimensions by Viazovska, work for which she was awarded the Fields Medal in 2022.

In physics, elliptic functions provide solutions to various boundary problems in electrostatics and fluid mechanics and play a fundamental role in the theory of integrable systems. Elliptic integrals arise in the solution to even the simplest mechanical problems, such as the pendulum.

The modular group $\mathrm{SL}(2,\mathbb{Z})$ first made its appearance in string theory in 1972 when Shapiro identified it as a symmetry of the integrand for the one-loop closed bosonic string amplitude. Shapiro defined the amplitude as the integral over the quotient of the Poincaré upper half plane by $\mathrm{SL}(2,\mathbb{Z})$ and argued that the amplitude thus obtained is free of the short-distance divergences that arise in quantum field theory. This fundamental observation extends to the perturbative superstring theories, and to all loop orders, provided that the group $\mathrm{SL}(2,\mathbb{Z})$ is replaced by the modular group $\mathrm{Sp}(2g,\mathbb{Z})$ for genus g Riemann surfaces. The absence of short-distance divergences uniquely qualifies string theory for the task of unifying gravity with the strong and electro-weak interactions into a consistent quantum theory.

A second context in which $\mathrm{SL}(2,\mathbb{Z})$ arises in string theory is as follows. The five perturbative superstring theories in ten-dimensional Minkowski spacetime with spacetime supersymmetry are the Type IIA and Type IIB theories with the maximal number of thirty-two supersymmetries, and the Type I and two Heterotic string theories with sixteen supersymmetries. The massless states of each one of these superstring theories are described by an associated supergravity theory, which is an extension of the Einstein–Hilbert theory of general relativity. For example, Type IIB supergravity contains, in addition to the space-time metric, a complex-valued axion-dilaton scalar field τ. The imaginary part of τ is related to the string coupling and must be positive on physical grounds. Thus, the field τ takes values in the Poincaré upper half plane $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$, where the group $\mathrm{SL}(2,\mathbb{R})$ acts on τ by Möbius transformations,

$$\tau \to \frac{a\tau + b}{c\tau + d} \qquad\qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{R})$$

While $\mathrm{SL}(2,\mathbb{R})$ is a symmetry of Type IIB supergravity, certain quantum effects in string theory reduce the $\mathrm{SL}(2,\mathbb{R})$ symmetry to its discrete subgroup $\mathrm{SL}(2,\mathbb{Z})$, also referred to as the part of the *S-duality group* of Type IIB string theory that acts on bosonic fields. Type IIB string solutions related by an $\mathrm{SL}(2,\mathbb{Z})$ transformation are physically identical, so that the space of inequivalent solutions is given by the double coset $\mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$. The implications of this conjectured symmetry were fully appreciated only with the discovery of NS- and D-brane solutions in the 1990s.

A third context where $\mathrm{SL}(2,\mathbb{Z})$ and higher arithmetic groups emerge is when string theory is considered on a space-time of the form $\mathbb{R}^{10-d}\times\mathbb{T}^d$, where $\mathbb{T}^d$ is a d-dimensional flat torus. Such a setup is referred to as *toroidal compactification*. While the Fourier analysis of supergravity fields on a torus produces only momentum modes, the compactification of a string theory produces both momentum and winding modes. The winding modes are responsible for quintessentially string-theoretic discrete symmetries, referred to as *T-dualities*, that have no counterpart in quantum field theory. For example, Type IIB superstring theory on a circle of radius R is T-dual to Type IIA superstring theory on a circle of radius α'/R, where α' is a constant with dimensions of length squared, which is proportional to the inverse of the string tension. Toroidal compactification converts some of the components of the bosonic fields, such as the metric, into scalar fields, which combine with the axion–dilaton field τ of Type IIB to live on a larger coset space that enjoys a larger arithmetic symmetry group. As a function of the dimension d, the following arithmetic symmetry groups $G(\mathbb{Z})$, and corresponding coset spaces $G(\mathbb{R})/K(\mathbb{R})$, arise starting with the ten-dimensional Type IIB superstring theory for $d=0$.

d	$G(\mathbb{Z})$	$G(\mathbb{R})$	$K(\mathbb{R})$
0	$\mathrm{SL}(2,\mathbb{Z})$	$\mathrm{SL}(2,\mathbb{R})$	$\mathrm{SO}(2)$
1	$\mathrm{SL}(2,\mathbb{Z})$	$\mathrm{SL}(2,\mathbb{R})\times\mathbb{R}^\times$	$\mathrm{SO}(2)$
2	$\mathrm{SL}(2,\mathbb{Z})\times\mathrm{SL}(3,\mathbb{Z})$	$\mathrm{SL}(2,\mathbb{R})\times\mathrm{SL}(3,\mathbb{R})$	$\mathrm{SO}(2)\times\mathrm{SO}(3)$
3	$\mathrm{SL}(5,\mathbb{Z})$	$\mathrm{SL}(5,\mathbb{R})$	$\mathrm{SO}(5)$
4	$\mathrm{SO}(5,5,\mathbb{Z})$	$\mathrm{SO}(5,5,\mathbb{R})$	$\big(\mathrm{SO}(5)\times\mathrm{SO}(5)\big)/\mathbb{Z}_2$
5	$E_{6,6}(\mathbb{Z})$	$E_{6,6}(\mathbb{R})$	$\mathrm{USp}(8)/\mathbb{Z}_2$
6	$E_{7,7}(\mathbb{Z})$	$E_{7,7}(\mathbb{R})$	$\mathrm{SU}(8)/\mathbb{Z}_2$
7	$E_{8,8}(\mathbb{Z})$	$E_{8,8}(\mathbb{R})$	$\mathrm{Spin}(16)/\mathbb{Z}_2$

On the last three lines, $E_{n,n}$ for $n=6,7,8$ denotes a particular real form of the corresponding complex Lie groups E_6, E_7, and E_8. Compactification on Calabi–Yau manifolds or orbifolds exhibit similarly quintessential string theoretic relations that go under the name of *mirror symmetry*.

A fourth context where the modular group $\mathrm{SL}(2,\mathbb{Z})$ arose in physics is

Yang–Mills theory. The Standard Model of Particle Physics is a Yang–Mills theory with gauge group $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$ in which the masses of quarks, leptons, and the gauge bosons $W^{\pm}, Z^0$ are generated via spontaneous symmetry breaking. In a *grand unified field theory*, the group $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$ itself arises by spontaneous symmetry breaking of a *simple gauge group* such as $\mathrm{SU}(5), \mathrm{SO}(10), E_6, E_7$, or E_8. The spectrum of these theories contains a massless photon and various electrically charged particles, but also contains ‘t Hooft–Polyakov magnetic monopoles. In 1977, Goddard, Nuyts, and Olive conjectured that such theories may exhibit *electric–magnetic duality* which swaps electric particles and magnetic monopoles. Dyons, which carry both electric and magnetic charges, extend this duality to the full modular group $\mathrm{SL}(2,\mathbb{Z})$, or a subgroup thereof. A concrete realization of electric–magnetic duality, referred to as Montonen–Olive duality, is provided by Yang–Mills theories with extended supersymmetry, and culminated in the Seiberg–Witten solution in 1994.

A fifth context where modular forms are of great importance is two-dimensional conformal field theory. Cardy linked the $\mathrm{SL}(2,\mathbb{Z})$ symmetry of a conformal field theory on a torus $\mathbb{T}^2$ to its unitarity properties. The powerful constraints on the operator product expansion of conformal primary fields in terms of representations of $\mathrm{SL}(2,\mathbb{Z})$, obtained by Erik Verlinde in 1988, may be implemented in the *modular bootstrap* program and have been used to great effect to advance the classification of conformal field theories. Thanks to the *gauge–gravity correspondence*, these constraints on conformal field theory give rise to constraints on theories of quantum gravity in three-dimensional anti-de Sitter space, which indeed was one of the original motivations for the modular bootstrap program. Additionally, different conformal field theories may be related to one another by Hecke operators.

Each of the contexts where modular forms play a role in physics, described in the previous paragraphs, continues to provide an active and fertile area of current research, with myriad open questions remaining. There are a number of further important and beautiful contexts where modular forms play a central role but that will not be addressed directly in this book. They include orbifold, Calabi–Yau, and F-theory compactifications; the counting of microstates of four-dimensional black holes; the various incarnations of moonshine; three-dimensional topological field theory; topological modular forms; and modular cosmology. No doubt, the theory of modular forms continues to teach us important conceptual and computational lessons about these well-weathered fields, to say nothing of its promise for up-and-coming topics such as generalized global symmetries.

The goal of this book is to exhibit the profound interrelations between modular forms and string theory, which are numerous. Our presentation is intended to be informal but mathematically precise, logically complete, and self-contained. We have made every effort to render the exposition as simple as possible and accessible to adventurous undergraduates, motivated graduate students, and dedicated professionals interested in the interface between theoretical physics and pure mathematics. To paraphrase Einstein, "Everything should be made as simple as possible, but not simpler."

Assuming little more than a knowledge of complex function theory, some planar differential geometry, and basic group theory, we introduce elliptic functions and elliptic curves as a lead-in to modular forms and modular curves for $SL(2,\mathbb{Z})$ and its congruence subgroups. A prior background in modular arithmetic, Riemann surfaces, or line bundles is not required as those subjects are presented in some detail in four separate appendixes. Free quantum fields on a torus provide an excellent illustration of how elliptic functions and modular forms can be used to solve problems in two-dimensional conformal field theory of relevance to string theory. A basic understanding of the operator formulation of quantum mechanics and some Lie algebra theory will prove useful here but is not absolutely required. As will be explained in the *organizational introductions*, further mathematical topics, ranging from quasi-modular forms and modular graph functions to Hecke operators and Galois theory, are included in order to broaden the spectrum of applications in conformal field theory and string theory.

Even the most economical introduction to string theory proper, attempted here, inevitably benefits from some familiarity with general relativity, classical fields, and some basic elements of scattering theory, though much may be picked up at an intuitive level during a first read-through. By contrast, the chapter on toroidal compactifications should be accessible without further physics prerequisites, while the chapter on S-duality will expose the reader to supergravity. Although the chapter on dualities in super Yang–Mills theory also appeals to several further physics concepts, such as Yang–Mills theory, supersymmetry, magnetic monopoles, and effective field theory, we have attempted to introduce each one of these vast subjects with the minimal amount of detail needed to exhibit their interplay with modular forms.

Part I

Modular forms and their variants

In Part I, we introduce the basic concepts and properties of elliptic functions, holomorphic and non-holomorphic modular forms, and a number of their variants that will be of interest to physics. Mostly mathematical aspects will be discussed here and little prior physics background is required.

In both mathematics and applications to physics, modular forms are intimately tied to elliptic functions. Thus, we start off Part I with a thorough exposition of elliptic functions, elliptic curves, Abelian differentials, and Abelian integrals in Chapter 2. Universal tools, such as Poisson resummation, the analytic continuation of parametric integrals, and the method of images are developed and illustrated along the way.

We introduce holomorphic and meromorphic modular forms under the modular group $\mathrm{SL}(2,\mathbb{Z})$ in Chapter 3, account for their properties, and construct them explicitly in terms of holomorphic Eisenstein series. Special attention is devoted to the discriminant cusp form Δ, the j-function, and the Dedekind η-function, which play key roles throughout.

In Chapter 4, we introduce variants of modular forms, including quasi-modular forms, almost-holomorphic modular forms, non-holomorphic Eisenstein series, Maass forms, mock modular forms, and quantum modular forms, and present the spectral decomposition theorem, all of which turn up in one physics application or another.

We illustrate the use of elliptic functions and modular forms in two-dimensional conformal field theory in Chapter 5 by constructing Green functions, correlation functions, and functional determinants for scalar and spinor fields on a torus and relating the results to the Kronecker limit formulas. The basics of conformal field theory are introduced here via various examples and no prior knowledge of the subject is required.

In Chapter 6, we review congruence subgroups of $\mathrm{SL}(2,\mathbb{Z})$ and their associated modular curves, elliptic points, cusps, and genera. Modular forms for the classic congruence subgroups are introduced in Chapter 7 and applied to counting the number of representations of an integer by sums of squares. In Chapter 8, we further generalize to vector-valued modular forms that transform under nontrivial representations of $\mathrm{SL}(2,\mathbb{Z})$, but whose individual components are modular forms under a congruence subgroup and systematically satisfy modular differential equations.

In Chapter 9, we introduce modular graph functions and forms, which provide nontrivial generalizations of non-holomorphic Eisenstein series, multiple zeta values, elliptic polylogarithms, and iterated modular integrals. Physically, they arise in the low energy expansion of superstring amplitudes, as will be shown in Chapter 12.

Cambridge University Press & Assessment
978-1-009-45753-8 — Modular Forms and String Theory
Eric D'Hoker , Justin Kaidi
Excerpt
More Information

2

Elliptic functions

In this chapter, elliptic functions are introduced via the method of images following a review of periodic functions, Poisson resummation, the unfolding trick, and analytic continuation applied to the Riemann ζ-function. The differential equations and addition formulas obeyed by periodic and elliptic functions are deduced from their series representation. The classic constructions of elliptic functions, in terms of their zeros and poles, are presented in terms of the Weierstrass $\wp$-function, the Jacobi elliptic functions $\mathrm{sn}, \mathrm{cn}, \mathrm{dn}$, and the Jacobi ϑ-functions. The elliptic function theory developed here is placed in the framework of elliptic curves, Abelian differentials, and Abelian integrals.

2.1 Periodic functions of a real variable

A function $f : \mathbb{R} \to \mathbb{C}$ is periodic with period 1 if for all $x \in \mathbb{R}$ it satisfies

$$f(x+1) = f(x) \tag{2.1.1}$$

Equivalently, f is a function of the circle $S^1 = \mathbb{R}/\mathbb{Z} \to \mathbb{C}$. The function f is completely specified by the values it takes in the unit interval $[0,1)$. There are various other useful systematic methods for constructing periodic functions, some of which we shall now review.

If a function $g : \mathbb{R} \to \mathbb{C}$ decays sufficiently rapidly at ∞, then a periodic function may be constructed by the *method of images*,

$$f(x) = \sum_{n \in \mathbb{Z}} g(x+n) \tag{2.1.2}$$

The functions $e^{2\pi i m x}$ for $m \in \mathbb{Z}$ are all periodic with period 1 and, in fact, form a basis for the square-integrable periodic functions or, equivalently, for $L^2(S^1)$. Their orthogonality and completeness are manifest from the following relations (in the sense of distributions),[1]

$$\int_0^1 dx\, e^{2\pi i m x}\, e^{-2\pi i m' x} = \delta_{m,m'}$$
$$\sum_{m\in\mathbb{Z}} e^{2\pi i m x}\, e^{-2\pi i m y} = \sum_{n\in\mathbb{Z}} \delta(x-y+n) \tag{2.1.3}$$

Any periodic function f with period 1 may be decomposed into a Fourier series as follows,

$$f(x) = \sum_{m\in\mathbb{Z}} f_m\, e^{2\pi i m x} \qquad\qquad f_m = \int_0^1 dx\, f(x)\, e^{-2\pi i m x} \tag{2.1.4}$$

The notation $e(x) = e^{2\pi i x}$ is often used in the mathematics literature, but we shall adhere to the physics notation and write the exponential explicitly.

2.1.1 Unfolding and Poisson resummation formulas

A very simple but extremely useful tool is the *unfolding trick*. If a function $g : \mathbb{R} \to \mathbb{C}$ decays sufficiently rapidly at ∞ to make its integral over $\mathbb{R}$ converge absolutely, then the unfolding trick gives,

$$\sum_{n\in\mathbb{Z}} \int_0^1 dx\, g(x+n) = \int_{-\infty}^{\infty} dx\, g(x) \tag{2.1.5}$$

More generally, combining the method of images with Fourier decomposition, we construct a periodic function f from a nonperiodic function g using (2.1.2) and then calculate the Fourier coefficients f_m using the unfolding trick (2.1.5),

$$f_m = \int_0^1 dx\, e^{-2\pi i m x} \sum_{n\in\mathbb{Z}} g(x+n) = \int_{-\infty}^{\infty} dx\, g(x)\, e^{-2\pi i m x} \tag{2.1.6}$$

The Fourier transform of g will be denoted by $\hat{g}$ and is given by,

$$\hat{g}(y) = \int_{-\infty}^{\infty} dx\, g(x)\, e^{-2\pi i x y} \tag{2.1.7}$$

so that $f_m = \hat{g}(m)$. Therefore, the function f may be expressed in two different ways,

$$f(x) = \sum_{n\in\mathbb{Z}} g(x+n) = \sum_{m\in\mathbb{Z}} \hat{g}(m)\, e^{2\pi i m x} \tag{2.1.8}$$

[1] Throughout, the Kronecker symbol $\delta_{m,n}$ equals 1 when $m = n$ and 0 otherwise, while the Dirac δ-function is normalized by $\int_{\mathbb{R}} dy\, \delta(x-y) f(y) = f(x)$ for an arbitrary test function $f(x)$.

Setting $x = 0$ (or any integer) gives the Poisson resummation formula.

Theorem 2.1.1. *A function g and its Fourier transform $\hat{g}$ defined in (2.1.7) obey the Poisson resummation formula,*

$$\sum_{n\in\mathbb{Z}} g(n) = \sum_{m\in\mathbb{Z}} \hat{g}(m) \tag{2.1.9}$$

assuming both sums are absolutely convergent.

An immediate application is to the case where g is a Gaussian. It will be convenient to normalize Gaussians as follows,

$$g(x) = e^{-\pi t x^2} \qquad\qquad \hat{g}(y) = \frac{1}{\sqrt{t}}\, e^{-\pi y^2/t} \tag{2.1.10}$$

The formulas in (2.1.10) are valid as long as $\mathrm{Re}\,(t) > 0$ and may be continued to $\mathrm{Re}\,(t) = 0$. The corresponding Poisson resummation formula then reads,

$$\sum_{n\in\mathbb{Z}} e^{-\pi t n^2} = \frac{1}{\sqrt{t}} \sum_{m\in\mathbb{Z}} e^{-\pi m^2/t} \tag{2.1.11}$$

We shall soon see that this relation admits an important generalization to Jacobi ϑ-functions, where it will correspond to a modular transformation.

2.1.2 Application to analytic continuation

The Riemann ζ-function, known already to Euler, is defined by the series,

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{2.1.12}$$

which converges absolutely for $\mathrm{Re}\,(s) > 1$ and thus defines a holomorphic function of s in that region. Its arithmetic significance derives from the fact that it admits a product representation, due again to Euler,

$$\zeta(s) = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1} \tag{2.1.13}$$

The Riemann ζ-function is just one example of the family of ζ-functions that may be associated with certain classes of self-adjoint operators whose spectrum is discrete, free of accumulation points, and bounded from below (we shall take them to be bounded from below by zero without loss of generality). Consider such a self-adjoint operator H and its associated discrete spectrum of real eigenvalues λ_n with $n \in \mathbb{N}$ and $\lambda_1 > 0$.[2] We may associate

[2] Throughout $\mathbb{N} = \{1, 2, 3, \dots\}$ denotes the set of positive integers.