3

Gold and brass: affine algebras and generalisations

This chapter introduces the nontwisted affine algebras – infinite-dimensional Lie algebras of considerable mathematical and physical interest – and searches for generalisations that preserve and enhance those special features. The affine algebras supply classic examples of Moonshine, in that the characters of their integrable modules are vector-valued Jacobi functions for $SL_2(\mathbb{Z})$. They thread through the remainder of the book, guiding all subsequent mathematical developments. Their Lie groups are discussed in Section 3.2.6.

Algebraically, the affine algebras naturally generalise to the Kac–Moody algebras (Section 3.3.1), although that generalisation seems to lose some of their magic. In turn, the Kac–Moody algebras generalise naturally to the Borcherds–Kac–Moody algebras (Section 3.3.2), which play a significant role in Borcherds' proof of Monstrous Moonshine through their denominator identities (Section 3.4.2). Two other natural generalisations of affine algebras are described elsewhere in Section 3.3. In Section 3.4.1 we study an important special case of what we later call the orbifold construction, and in the final subsection we touch on a more recent and tangential development.

The Virasoro algebra (Section 3.1.2) plays a prominent structural role in conformal field theory (Chapter 4) and vertex operator algebras (Chapter 5); its relation to moduli spaces is a fundamental source of Moonshine itself.

3.1 Modularity from the circle

3.1.1 Central extensions

Let V be any (complex) vector space, and let $\mathrm{GL}(V)$ denote the group of all invertible linear maps $V \to V$. A *projective representation* of a group G is a map $P : G \to \mathrm{GL}(V)$ such that $P(e) = I$ (the identity), and given any elements $g, h \in G$, there is a nonzero complex number $\alpha(g, h)$ such that

$$P(g)\,P(h) = \alpha(g, h)\,P(gh). \qquad (3.1.1a)$$

We call P an *α-representation*. So just as a (true) representation is a group homomorphism $R : G \to \mathrm{GL}(V)$, a projective representation defines a group homomorphism P from G into the projective group $\mathrm{PGL}(V) := \mathrm{GL}(V)/\{\mathbb{C}^\times I\}$ (hence the name); conversely, given a homomorphism $\pi : G \to \mathrm{PGL}(V)$, arbitrarily choosing a 'section', that is a representative $P(g) \in \mathrm{GL}(V)$ in each equivalence class $\pi(g) \in \mathrm{PGL}(V)$, defines a projective representation of G. A projective representation P is a true representation iff $\alpha(g, h) = 1$ for all $g, h \in G$.

Projective representations are plentiful. For example, the multiplier μ in Definition 2.2.1 is a projective representation of $\mathrm{SL}_2(\mathbb{Z})$ whenever the weight k is rational. Quasi-periodicity (2.3.5a) is a projective representation of the abelian group $\mathbb{C}^2$ on the space of functions $f : \mathbb{C} \to \mathbb{C}$. In quantum physics (Section 4.2) the state of a system is completely described by a nonzero vector v in a Hilbert space. However, any nonzero multiple λv describes a physically identical state. Thus projective representations arise naturally also in quantum physics, where they are called 'ray representations'.

Note that associativity

$$\begin{aligned}\alpha(h,k)\,\alpha(g,hk)\,P(ghk) &= P(g)\,(P(h)\,P(k)) = (P(g)\,P(h))\,P(k)\\ &= \alpha(g,h)\,\alpha(gh,k)\,P(ghk)\end{aligned}$$

tells us that

$$\alpha(h,k)\,\alpha(g,hk) = \alpha(gh,k)\,\alpha(g,h), \qquad \forall g,h,k \in G. \tag{3.1.1b}$$

This equation may remind the reader of a two-cocycle condition, hinting of the relevance of cohomology. Indeed, this function $\alpha : G \times G \to \mathbb{C}^\times$ is called a *2-cocycle* and group cohomology organises the projective representations.

Two projective representations $P_i : G \to \mathrm{GL}(V_i)$ are *(linearly) equivalent* if there is a vector space isomorphism $\varphi : V_1 \to V_2$ such that $\varphi^{-1} \circ P_1 \circ \varphi = P_2$. Equivalent projective representations must have the same 2-cocycle α. For a given α, the number of inequivalent irreducible α-representations of G equals the number of conjugacy classes of α-regular elements $g \in G$ (g is called *α-regular* if $\alpha(g, h) = \alpha(h, g)$ for all $h \in C_G(g)$). Hence this number is at most the number of inequivalent irreducible true G-representations.

We call projective representations $P_i : G \to \mathrm{GL}(V_i)$ *projectively equivalent* when there is a function $\beta : G \to \mathbb{C}^\times$ and a vector space isomorphism $\varphi : V_1 \to V_2$ such that

$$\varphi^{-1}(P_1(\varphi(g))) = \beta(g)\,P_2(g), \qquad \forall g \in G.$$

The 2-cocycles of projectively equivalent projective representations are related by

$$\alpha_2(g,h) = \alpha_1(g,h)\,\beta(gh)\,\beta^{-1}(g)\,\beta^{-1}(h).$$

β plays the role of a coboundary, so the 2-cocycles α_i of projectively equivalent projective representations lie in the same cohomology class $[\alpha] \in H^2(G, \mathbb{C}^\times)$, and $H^2(G, \mathbb{C}^\times)$ classifies the projectively inequivalent projective representations. $H^2(G, \mathbb{C}^\times)$ is an abelian group, called the *Schur multiplier*, and is finite when G is finite. The point of converting a problem into algebraic topology is that machinery (and experts!) are available to help compute these groups. For example, $H^2(\mathbb{Z}_n, \mathbb{C}^\times) = H^2(\mathrm{SL}_2(\mathbb{Z}), \mathbb{C}^\times) = H^2(\mathbb{M}, \mathbb{C}^\times) = \{0\}$ while $H^2(Co_1, \mathbb{C}^\times) \cong \mathbb{Z}_2$. This implies, for instance, that any projective representation of the Monster $\mathbb{M}$ is projectively equivalent to a true representation of $\mathbb{M}$.

Projective representations of Lie algebras are defined similarly: $P : \mathfrak{g} \to \mathrm{End}(V)$ is linear, and equations (3.1.1) become

$$[P(x), P(y)] = P([x, y]) + c(x, y)\, I, \tag{3.1.2a}$$

$$c(x, y) = -c(y, x), \tag{3.1.2b}$$

$$c([xy], z) = c([yz], x) + c([zx], y) = 0, \tag{3.1.2c}$$

where the 2-cocycle c is complex-valued and I is the identity endomorphism.

Geometrically, projective representations often arise from the following fundamental construction. Let $\mathcal{L} \to M$ be any line bundle with connection ∇ over some manifold M (Section 1.2.2). Let $\varphi : \mathfrak{g} \to \mathrm{Vect}(M)$ be a homomorphism from some Lie algebra $\mathfrak{g}$ to the Lie algebra of vector fields on M. The map $x \mapsto \nabla_{\varphi(x)}$, sending $x \in \mathfrak{g}$ to the covariant derivative in the direction $\varphi(x)$, associates with each $x \in \mathfrak{g}$ a differential operator on the space of sections of $\mathcal{L}$. Since

$$[\nabla_X, \nabla_Y] = \nabla_{[X,Y]} + R(X, Y)\, I$$

for each vector field X, Y, where R is the curvature of the connection, this map defines a projective representation of $\mathfrak{g}$ on the space $\Gamma(\mathcal{L})$ of sections of $\mathcal{L}$, with cocycle $c = R$. As we will see later this chapter, the central extensions of both the Witt and the loop algebras can be interpreted in this way [**13**]. This construction is well known in physics, where it falls under the slogan ‘curvature is a local anomaly’ (by contrast, global anomalies are monodromy effects like modularity).

A standard trick (central extensions) converts *projective* representations into *true* representations. Let G be any group, and let A be any abelian group. By a *central extension* $\widehat{G}$ of G by A, we mean that A can be identified with a subgroup of the centre of $\widehat{G}$, and the quotient $\widehat{G}/A$ is isomorphic to G. For example, the dihedral group $\mathcal{D}_4$ is a central extension (by $\mathbb{Z}_2$) of a central extension (by $\mathbb{Z}_2$) of a central extension (by $\mathbb{Z}_2$) of $\{e\}$.

Let P be a projective representation of a group G, and assume for simplicity that no operator $P(g)$ is a scalar multiple $a\,I$ of the identity. Let $\widehat{G}$ be the group consisting of all operators $a\,P(g)$, for $a \in \mathbb{C}^\times$ and $g \in G$. Then $\widehat{G}$ is a central extension of G by $\mathbb{C}^\times$, and $\widehat{G}$ is defined by a faithful representation in V. The projective representation of G has been transformed into a true representation of the larger group $\widehat{G}$. The specific situation for finite groups and the most common finite-dimensional Lie groups is simpler:

Theorem 3.1.1 (a) *Let G be a finite group. Then there is a central extension $\widetilde{G}$ of G by its Schur multiplier $H^2(G, \mathbb{C}^\times)$, with the following property: any projective representation $P : G \to \mathrm{GL}(V)$ of G lifts to a true representation $\widetilde{P} : \widetilde{G} \to \mathrm{GL}(V)$ of $\widetilde{G}$.*
(b) *Let G be a connected, finite-dimensional semi-simple Lie group over $\mathbb{R}$ or $\mathbb{C}$, and let $\widetilde{G}$ be its universal cover group (which is a central extension of G by the fundamental group $\pi_1(G)$). Then any continuous finite-dimensional projective representation $P : G \to \mathrm{GL}(V)$ of G lifts to a true representation $\widetilde{P} : \widetilde{G} \to \mathrm{GL}(V)$ of $\widetilde{G}$.*

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

Conversely, a true representation of $\widetilde{G}$ restricts to a projective representation of G. The central extension $\widetilde{G}$ in Theorem 3.1.1(a) is a finite group (e.g. for $\widetilde{Co_1}$ take the Conway group Co_0), and in (b) is a Lie group of the same dimension as G (see Theorem 1.4.3). For Lie groups there is a topological (π_1) as well as cohomological (H^2) obstacle to the trivialisation of projective representations. The assumption in (b) that G be semi-simple was made only to guarantee that the Schur multiplier of G would be trivial. The conclusion to Theorem 3.1.1(b) also holds for certain non-semi-simple Lie groups, such as the Poincaré group important to relativistic physics. On the other hand, the Galilei group, which plays the same role in pre-relativistic physics, has nontrivial Schur multiplier. In this case, the relevant cover will be a Lie group of higher dimension. The simplest example of this phenomenon is the additive group $\mathbb{C}^2$, and its central extension the three-dimensional Heisenberg group (see Question 3.1.3). It is through projective representations of $\mathbb{C}^2$ that the Heisenberg group and algebra arise in both theta functions (Section 2.4.2) and quantum physics (Section 4.2). Similarly, the Galilei group must act on nonrelativistic wave-functions (i.e. solutions to the Schrödinger equation (4.2.1)) *projectively* – this is a consequence of the nontriviality of the Schur multiplier of the Galilei group (Question 4.2.1).

Incidentally, the Schur multiplier $H^2(G, \mathbb{C}^\times)$ of a finite group G appears in another context. Consider any presentation of G, with say m generators and n relations. The finiteness of G requires that $m \leq n$. The Schur multiplier of G is a finite abelian group, so let h be its number of generators as in Theorem 1.1.1. Then $n - m \geq h$.

We are primarily interested in *one-dimensional central extensions* $\widehat{\mathfrak{g}}$ of Lie algebras $\mathfrak{g}$, that is a vector space $\widehat{\mathfrak{g}} = \mathfrak{g} \oplus \mathbb{C}C$ together with the brackets

$$[ab]_{\text{new}} = [ab]_{\text{old}} + c(a,b)\,C, \tag{3.1.3a}$$

$$[aC] = 0. \tag{3.1.3b}$$

The element C is called the *central term*. Equivalently, we have

$$0 \to \mathbb{C} \to \widehat{\mathfrak{g}} \to \mathfrak{g} \to 0, \tag{3.1.3c}$$

together with the requirement that the image $\mathbb{C}C$ of $\mathbb{C}$ in $\widehat{\mathfrak{g}}$ is in the centre of $\widehat{\mathfrak{g}}$. The short exact sequence (3.1.3c) says that there is an ideal in $\widehat{\mathfrak{g}}$ (namely the image of the second arrow) isomorphic as a Lie algebra to $\mathbb{C}$, and that when this ideal is projected out (by the third arrow) we recover $\mathfrak{g}$.

The exact sequence (3.1.3c) has the charm of not requiring an explicit *splitting* of $\widehat{\mathfrak{g}}$ into a $\mathfrak{g}$-part $\overline{\mathfrak{g}}$ (namely, a *lift* of the Lie algebra $\mathfrak{g}$ onto a subspace $\overline{\mathfrak{g}}$) and a $\mathbb{C}$-part $\mathbb{C}C$. The point is that there are many possible splittings: for example, given any such splitting $\widehat{\mathfrak{g}} = \overline{\mathfrak{g}} \oplus \mathbb{C}C$, choose a linear map $f : \overline{\mathfrak{g}} \to \mathbb{C}$; then a new splitting is obtained by replacing the subspace $\overline{\mathfrak{g}}$ with the span of the $a + f(a)C$, as a runs through $\overline{\mathfrak{g}}$. Modern mathematics abhors arbitrary choices, and so would encourage us to delay the choice of such a splitting as long as Good Fortune permits. Of course this is merely the current century-long fad, and there are advantages and disadvantages to it, and indeed physics prefers the opposite choice.

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

$\widehat{\mathfrak{g}}$ will be a Lie algebra iff the function $c : \mathfrak{g} \times \mathfrak{g} \to \mathbb{C}$ obeys (3.1.2b), (3.1.2c); as before, c is called the *2-cocycle* associated with the extension (3.1.3). The trivial 2-cocycle $c \equiv 0$ always works, in which case $\widehat{\mathfrak{g}}$ is merely the Lie algebra direct sum $\mathfrak{g} \oplus \mathbb{C}$.

We regard two extensions $\widehat{\mathfrak{g}}_1, \widehat{\mathfrak{g}}_2$ as equivalent if there is a Lie algebra isomorphism $\varphi : \widehat{\mathfrak{g}}_1 \to \widehat{\mathfrak{g}}_2$ that sends the ideal $\mathbb{C}C_1$ of $\widehat{\mathfrak{g}}_1$ onto $\mathbb{C}C_2 \subset \widehat{\mathfrak{g}}_2$. One way (but not the only way) to get equivalent extensions is to change the splitting $\widehat{\mathfrak{g}} = \overline{\mathfrak{g}} \oplus \mathbb{C}C$, as mentioned before. In the language of Lie algebra cohomology (see e.g. [**183**] for a mathematical treatment, or [**27**] for a physically motivated one), $f : \overline{\mathfrak{g}} \to \mathbb{C}$ is a 2-coboundary, and the resulting 2-cocycles c_1, c_2 define the same class in the cohomology space $H^2(\mathfrak{g})$. There are other ways though to obtain equivalent extensions – for example, the central term can be rescaled – so $H^2(\mathfrak{g})$ is in general too fine to serve as a 'moduli space' of one-dimensional central extensions of $\mathfrak{g}$, but it gives a very useful partial answer. For example, $H^2(\mathfrak{g})$ is trivial for any finite-dimensional semi-simple Lie algebra $\mathfrak{g}$, which means any such $\mathfrak{g}$ has only trivial central extensions (see Question 3.1.4).

For a concrete example, consider the n-dimensional abelian Lie algebra $\mathfrak{h} = \mathbb{C}^n$, with basis $\{e_1, \ldots, e_n\}$. A one-dimensional central extension $\widehat{\mathfrak{h}}$ of $\mathfrak{h}$ is uniquely determined by n^2 numbers $\alpha_{ij} \in \mathbb{C}$ defined by $[e_i e_j] = \alpha_{ij} C$, where $C \in \widehat{\mathfrak{h}}$ is central (all other brackets of $\widehat{h}$ are determined by bilinearity and $[e_i C] = 0$). Anti-commutativity requires $\alpha_{ij} = -\alpha_{ji}$, and anti-associativity is automatically satisfied. Thus each choice of an anti-symmetric $n \times n$ matrix $A = (\alpha_{ij})$ defines a one-dimensional central extension $\widehat{\mathfrak{h}}_A$ of $\mathfrak{h} = \mathbb{C}^n$, and conversely. The dependence of this argument on an arbitrary choice of basis e_i means there is redundancy here: in particular, two such central extensions $\widehat{\mathfrak{h}}_A$ and $\widehat{\mathfrak{h}}_{B'}$ define isomorphic Lie algebras iff there is an invertible matrix B such that $A' = BAB^t$. The reader can verify that any anti-symmetric matrix A is equivalent in this sense to the direct sum of k copies of $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, and $\ell = n - 2k$ copies of (0), where $2k$ is the rank of A. Thus we get a different one-dimensional central extension of $\mathbb{C}^n$, for each $k = 0, 1, \ldots, \lfloor n/2 \rfloor$. When A is invertible (i.e. $k = n/2$), we call $\widehat{\mathfrak{h}}$ a Heisenberg algebra; as simple a (non-simple!) Lie algebra as it is, it's one of the most important.

3.1.2 The Virasoro algebra

Recall the Witt algebra $\mathfrak{Witt}$ in (1.4.9). For each choice of $\alpha, \beta \in \mathbb{C}$, we get a module $V_{\alpha,\beta}$, with basis v_k, $k \in \mathbb{Z}$, given by

$$\ell_n.v_k = -(k + \alpha + \beta + \beta n)\, v_{k+n}. \tag{3.1.4a}$$

This can be obtained from the derived module (Section 1.5.5) coming from the natural action of a subgroup of the diffeomorphism group Diff(S^1) on the space of differential 'forms' $p(z)\, z^\alpha\, (\mathrm{d}z)^\beta$, where $p(z) \in \mathbb{C}[z^{\pm 1}]$ are Laurent polynomials. Clearly, $V_{\alpha+m,\beta} \cong V_{\alpha,\beta}$ for any $m \in \mathbb{Z}$.

As usual, we are interested in unitary modules (Section 1.5.1), and for this we need an anti-homomorphism ω of $\mathfrak{Witt}$. Up to an automorphism of $\mathfrak{Witt}$, the unique choice

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

is $\omega \ell_n = \ell_{-n}$. Then for this choice, $V_{\alpha,\beta}$ is unitary iff both $\mathrm{Re}(\beta) = 1/2$ and $\alpha + \beta \in \mathbb{R}$ [**334**]. These modules are also irreducible.

The element $\ell_0 \in \mathfrak{Witt}$ is obviously special and plays the role of energy operator (Hamiltonian) in the application to physics. The most interesting $\mathfrak{Witt}$-modules are unitary ones with diagonalisable ℓ_0. In this case the eigenvalues of ℓ_0 will necessarily be real, and should have the physical interpretation of energy. Unfortunately, the only nontrivial unitary irreducible $\mathfrak{Witt}$-modules with ℓ_0 diagonalisable are those $V_{\alpha\beta}$. This is unfortunate because the eigenvalues of ℓ_0 in any $V_{\alpha\beta}$ have no upper or lower bound. For reasons of stability, physics wants energy to be bounded below. The space $V_{\alpha\beta}$ is infinite-dimensional, but ℓ_0 defines on it a natural grading into finite-dimensional subspaces, and so we are led to formally define its *graded-dimension* to be

$$\mathrm{tr}_{V_{\alpha\beta}} q^{\ell_0} = \sum_{k \in \mathbb{Z}} q^{k+\alpha+\beta}. \tag{3.1.4b}$$

Unfortunately this never converges.

Central extensions are a common theme in infinite-dimensional Lie theory.[1] Their *raison d'être* is always the same: a richer supply of representations. The *Virasoro algebra* $\mathfrak{Vir}$ is the one-dimensional central extension $\mathfrak{Vir} = \mathfrak{Witt} \oplus \mathbb{C}C$ with brackets

$$[L_m L_n] = (m-n)L_{m+n} + \delta_{n,-m} \frac{m(m^2-1)}{12} C, \tag{3.1.5a}$$

$$[L_m C] = 0. \tag{3.1.5b}$$

As always, we avoid convergence issues by defining $\mathfrak{Vir}$ to consist of only finite linear combinations of these basis vectors. Incidentally, a common mistake in the physics literature is to regard C as a number: it is in fact a vector, though in most modules of interest to, for example, mathematical physics it is mapped to a scalar multiple cI of the identity.

The reason for the strange-looking (3.1.5a) is that we have little choice: $\mathfrak{Vir}$ is the unique nontrivial one-dimensional central extension of $\mathfrak{Witt}$ (Question 3.1.5). The factor $\frac{1}{12}$ there is conventional, but arises naturally in the realisations of $\mathfrak{Vir}$ by normal-ordered operators in Fock spaces (see (3.2.13), (3.2.14) for such a calculation). In fact, the normal-ordering prescription is somewhat arbitrary and actually we are much more interested in a slightly different basis of $\mathfrak{Vir}$, with L_0 replaced by $L_0 - C/24$. This is the combination appearing in almost every expression for characters from this point on. Where does this $-C/24$ come from? With this modified L_0, the brackets (3.1.5a) simplify (Question 3.1.8). According to conformal field theory or vertex operator algebras, this new basis corresponds to a change in topology (see Section 5.3.4), which can be calculated using the Atiyah–Singer Index Theorem [**8**], so physically the 'conformal anomaly' term $-c/24$ is a Casimir effect. But the best algebraic explanation for this $-c/24$ is given Section 3.2.3.

As before, $L_0 \in \mathfrak{Vir}$ is the energy operator, and so we want irreducible $\mathfrak{Vir}$-modules where L_0 is diagonalisable and its eigenvalues are bounded below. Let v be any eigenvector of L_0 in such a module, say $L_0 v = Ev$, and suppose $L_n v \neq 0$ for some $n > 0$. Then

[1] On the other hand, the *finite-dimensional* simple Lie algebras do not have nontrivial central extensions.

$L_0(L_n v) = -nL_n v + L_n L_0 v = (E - n)L_n v$ and thus $(L_n)^\ell v$ will be an eigenvector of L_0 whose eigenvalue $E - n\ell$ has real part going to $-\infty$ as $\ell \to \infty$. Thus any $\mathfrak{Vir}$-module whose L_0-eigenvalues have real part bounded below must be a *highest-weight module*.

More precisely, because $\mathfrak{Vir}$ has a triangular decomposition (recall (1.5.5d))

$$\mathfrak{Vir}_- \oplus \mathfrak{Vir}_0 \oplus \mathfrak{Vir}_+ = \text{span}\{L_n\}_{n<0} \oplus \text{span}\{L_0, C\} \oplus \text{span}\{L_n\}_{n>0},$$

we can mimic the construction of highest-weight modules in Section 1.5.3. In particular, for any $h, c \in \mathbb{C}$, the *Verma module* $M(c, h)$ is the universal $\mathfrak{Vir}$-module generated by a vector $v \neq 0$ obeying

$$L_0 v = hv,\ Cv = cv,\ L_n v = 0, \qquad \forall n > 0.$$

The pair (c, h) is the highest weight; c is the *central charge* and h the *conformal weight*. As before, it can be more explicitly defined using the universal enveloping algebra, or equivalently by inducing the module from $\mathfrak{Vir}_0 \oplus \mathfrak{Vir}_+$ to all of $\mathfrak{Vir}$. By the Poincaré–Birkhoff–Witt Theorem 1.5.2, $M(c, h)$ has a basis given by all vectors

$$L_{-i_1} L_{-i_2} \cdots L_{-i_n} v,$$

for all integers $i_1 \geq i_2 \geq \cdots \geq i_n \geq 1$. Any other $\mathfrak{Vir}$-module with highest weight (c, h) is a homomorphic image of $M(c, h)$, or equivalently the quotient of $M(c, h)$ by some ideal.

Each Verma module $M(c, h)$ is indecomposable, but may not be irreducible. However, they all have a unique nontrivial irreducible quotient $V(c, h)$, which is then the unique irreducible $\mathfrak{Vir}$-module with highest weight (c, h).

The anti-linear anti-homomorphism ('adjoint') of $\mathfrak{Vir}$ sends L_n to L_{-n}, and fixes C. The only unitary irreducible $\mathfrak{Vir}$-modules where L_0 is diagonalisable and all its eigenspaces are finite-dimensional are certain $V_{\alpha,\beta}$ in (3.1.4a) (these are $\mathfrak{Vir}$-modules with C acting trivially), as well as certain highest-weight modules $V(c, h)$ and their duals, the lowest-weight modules $V(c, h)^\star$. In fact, $V(c, h)$ (and $V(c, h)^\star$) are unitary iff either: (i) both $c \geq 1$ and $h \geq 0$; or (ii) c and h fall into the *discrete series*, i.e. for $m, r, s \in \mathbb{N}$ with $1 \leq s \leq r \leq m + 1$,

$$c = c_m := 1 - \frac{6}{(m+2)(m+3)},\ h = h_{m;rs} := \frac{((m+3)r - (m+2)s)^2 - 1}{4(m+2)(m+3)}. \quad (3.1.6)$$

These $V(c, h)$ are called *positive-energy representations* since the spectrum of L_0 is positive. Thus the only unitary irreducible $\mathfrak{Vir}$-modules with L_0 diagonalisable, with finite-dimensional L_0-eigenspaces, and with the L_0-spectrum bounded below, are the $V(c, h)$ in (i) and (ii). They are the building blocks of the most interesting affine algebra representations, vertex operator algebra modules and conformal field theories.

For unitary $V(c, h)$, we have $V(c, h) = M(c, h)$ when both $c > 1$ and $h > 0$, or when $c = 1$ and $2\sqrt{h} \notin \mathbb{Z}$. In these cases, by analogy with (3.1.4b), $V(c, h)$ has

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

graded-dimension

$$\dim_{V(c,h)}(q) := \mathrm{tr}_{V(c,h)} q^{L_0} = q^h \prod_{n=1}^{\infty}(1-q^n)^{-1}, \tag{3.1.7a}$$

as the infinite product gives the generating function for the partition numbers:

$$\prod_{n=1}^{\infty}(1-q^n)^{-1} = \sum_{m=1}^{\infty} p(m) q^m \tag{3.1.7b}$$

where $p(m)$ is the number of ways to write m as a sum $m = a_1 + a_2 + \cdots + a_k$ for positive integers $1 \le a_1 \le a_2 \le \cdots \le a_k$. Unlike (3.1.4b), this converges whenever $|q| < 1$. In fact, we recognise (3.1.7b) as (up to a factor of $q^{1/24}$) the reciprocal of the Dedekind eta $\eta(\tau)$ (2.2.6b), once we change variables by $q = e^{2\pi \mathrm{i}\tau}$ – we saw last chapter that $\eta(\tau)$ is a modular form for $\mathrm{SL}_2(\mathbb{Z})$. In fact we obtain

$$\dim_{V(c,h)}\left(e^{2\pi \mathrm{i}(\tau+1)}\right) = e^{2\pi \mathrm{i}(h-\frac{1}{24})} \dim_{V(c,h)}(e^{2\pi \mathrm{i}\tau}), \tag{3.1.7c}$$

$$\dim_{V(c,h)} e^{-2\pi \mathrm{i}/\tau} = \sqrt{\frac{\mathrm{i}}{\tau}} \int_{-\infty}^{\infty} \exp[2\pi \mathrm{i} h h'] \dim_{V(c,h)} e^{2\pi \mathrm{i}\tau}\, \mathrm{d}h'. \tag{3.1.7d}$$

This is our first glimpse of modularity from a graded dimension, though it certainly won't be our last. But $\eta(\tau)$ arises here through elementary combinatorics, so it is tempting to dismiss this modularity as accidental. This however would be an error.

What should be the characters of these $\mathfrak{Vir}$-modules? For simple Lie algebras, we define the character as a trace over formal exponentials of elements of the Cartan subalgebra. The analogue of the Cartan subalgebra here is $\mathfrak{Vir}_0 = \mathbb{C}L_0 \oplus \mathbb{C}\,C$, so the character of $V(c,h)$ should be

$$\mathrm{ch}_{c,h}(z_L, z_C) := \mathrm{tr}_{V(c,h)} e^{2\pi \mathrm{i} z_L L_0 + 2\pi \mathrm{i} z_C C}, \tag{3.1.8}$$

which equals $e^{2\pi \mathrm{i} c z_C}$ times the graded-dimension of $V(c,h)$ (with $q = e^{2\pi \mathrm{i} z_L}$).

The characters of the discrete series (3.1.6) are calculated in [**477**], and again converge for $|e^{2\pi \mathrm{i} z_L}| < 1$. Moreover, they obey a much more interesting modularity than do the graded-dimensions in (3.1.7): let $\begin{pmatrix} a & b \\ f & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ act on $\mathfrak{Vir}_0$ by

$$(z_L, z_C) \mapsto \left(\frac{a z_L + b}{f z_L + d}, z_C + \frac{f z_L^2 + (d-a) z_L - b}{24\,(f z_L + d)}\right); \tag{3.1.9}$$

then $\mathrm{ch}_{c_m, h_{m;rs}}(z_L, z_C)$ is fixed by some $\Gamma(N)$ (recall (2.2.4a)), and for each fixed m (i.e. fixed central charge c), the span over all $1 \le s \le r \le m+1$ of the characters $\mathrm{ch}_{c_m, h_{m,rs}}$ is invariant under $\mathrm{SL}_2(\mathbb{Z})$. They furnish a good example of *modular data* (Definition 6.1.6). This $\mathrm{SL}_2(\mathbb{Z})$ action (3.1.9) is a little complicated; if instead we specialise to the variables $z_L = \tau$ and $z_C = -\tau/24$, then each

$$\mathrm{ch}_{c_m; h_{m;rs}}(\tau) := \mathrm{ch}_{c_m; h_{m;rs}}(\tau, -\tau/24) = e^{-2\pi \mathrm{i}\, c/24}\, \mathrm{tr}_{V(c,h)} e^{2\pi \mathrm{i} z_L L_0} \tag{3.1.10}$$

is a modular function for some $\Gamma(N)$ for $\tau \in \mathbb{H}$, and for fixed m the characters $\mathrm{ch}_{c_m; h_{m;rs}}(\tau)$ form a vector-valued modular function for $\mathrm{SL}_2(\mathbb{Z})$ (Definition 2.2.2).

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

The best explanation for the mysterious-looking discrete series (3.1.6) will probably come from the orbit method [**563**], but the analysis is still incomplete. At least part of the discrete series of the Virasoro algebra has been related to (co)homology theory of the universal cover of $\mathrm{SL}_2(\mathbb{R})$, given a discrete topology [**164**]. This should be explored further.

The characters of the non-unitary $V(c, h)$, for $c, h \in \mathbb{R}$, have most of the properties of those of the unitary ones, and it is unfair to completely ignore them. For example, for $c, h \in \mathbb{R}$ the modules $V(c, h)$ have a contravariant nondegenerate Hermitian form $\langle\star, \star\rangle$, apart from the positive-definiteness condition. Lie algebras typically have too many representations and some criterion is needed that isolates the interesting ones, but unitarity is too restrictive here.

As we know, the Lie algebra $\mathrm{Vect}(S^1)$ of vector fields on the circle contains the real Witt algebra $\mathfrak{Witt}_{\mathbb{R}}$ (i.e. the span over $\mathbb{R}$ of the generators ℓ_n in (1.4.9)) as a dense 'Laurent polynomial' subalgebra. The connected real Lie group naturally associated with $\mathrm{Vect}(S^1)$ is the group $\mathrm{Diff}^+(S^1)$ of orientation-preserving diffeomorphisms $S^1 \to S^1$ of the circle. As a group, $\mathrm{Diff}^+(S^1)$ is simple [**286**] but as a manifold it is not simply connected: its universal cover $\widetilde{\mathrm{Diff}}(S^1)$ is the group of all diffeomorphisms $\phi : \mathbb{R} \to \mathbb{R}$ of the real line satisfying the periodicity condition $\phi(x + 2\pi) = \phi(x) + 2\pi$. The centre of the universal cover is $\mathbb{Z}$ (namely $\phi_n(x) = x + 2\pi n$) and $\widetilde{\mathrm{Diff}}(S^1)/\mathbb{Z} \cong \mathrm{Diff}^+(S^1)$.

Nontrivial central extensions of $\mathrm{Diff}^+(S^1)$ by a circle are explicitly constructed in, for example, section 6.8 of [**465**] and appendix D.5 of [**295**]; these all have a Lie algebra isomorphic to the real Virasoro algebra $\mathfrak{Vir}_{\mathbb{R}}$ (i.e. the $\mathbb{R}$-span of the generators L_m, C of (3.1.5)).

Lie theory for the Virasoro and Witt algebras (and more generally the Lie algebra $\mathrm{Vect}(M)$ of vector fields on any manifold M) is much more complicated than the finite-dimensional semi-simple theory described in Chapter 1. For example, although the 'exponential' map exp: $\mathrm{Vect}(S^1) \to \mathrm{Diff}^+(S^1)$ is defined here (by first integrating the vector field to its flow), it is neither locally one-to-one nor locally onto (proposition 3.3.1 of [**465**]). By comparison, the exponential map of compact Lie groups is locally one-to-one and globally onto. Moreover, the complex Lie algebra $\mathbb{C} \otimes \mathrm{Vect}(S^1)$ does not have a corresponding Lie group. After all, although a vector field on S^1 corresponds to a path in the space of maps (in fact diffeomorphisms) $S^1 \to S^1$, and these form a group by composition, a *complex* vector field on S^1 corresponds to a path in the space of maps $S^1 \to \mathbb{C}$ and these won't form a group. Segal [**502**] suggests that the complex Lie semigroup $\mathbf{C}_{0,2}$ defined in Section 4.4.1 is the closest we can come to the complexification of $\mathrm{Diff}^+(S^1)$.

We have two fairly general frameworks in which to understand Lie group representations: Borel–Weil and the orbit method (*a.k.a.* geometric quantisation). There is, as we recall from Section 1.5.5, a general philosophy that says the representations of a group G (here $\mathrm{Diff}(S^1)$) are in one-to-one correspondence with certain orbits of the coadjoint action of G on the Lie algebra $\mathfrak{g}$ of G (here $\mathfrak{Witt}$). As mentioned earlier, Witten [**563**] explored this possible relation for the Virasoro algebra. For example, the homogeneous space $\mathrm{Diff}(S^1)/S^1$ appears as an orbit, and can be associated with ghosts in string theory.

The main motivation would be to find a new interpretation for the discrete series (3.1.6), which is a little mysterious from the algebraic point of view. Witten identified the orbits to which these should correspond, but couldn't quantise those orbits (this is a common curse of the orbit method).

The space $\mathrm{Diff}(S^1)/\mathrm{PSL}_2(\mathbb{R})$ is also a coadjoint orbit. Something special happens here when we replace $\mathrm{Diff}(S^1)$ with the larger group $\mathrm{QS}(S^1)$ of *quasi-symmetric* homeomorphisms of S^1: then $\mathrm{QS}(S^1)/\mathrm{PSL}_2(\mathbb{R})$ is called the *universal Teichmüller space* $\mathfrak{T}$. Every Teichmüller space $\mathfrak{T}_{g,n}$ (recall Section 2.1.4) is naturally contained in $\mathfrak{T}$. Likewise, $\mathrm{Diff}(S^1)/\mathrm{PSL}_2(\mathbb{R})$ naturally embeds in $\mathfrak{T}$ (every diffeomorphism of S^1 is quasi-symmetric), and intersects each $\mathfrak{T}_{g,n}$ transversely. See the reviews [**460**], [**168**] for definitions and references. Given this, an intriguing answer to the challenge suggested by Manin in Section 5.4.1 is to consider the reparametrisations of strings using quasi-symmetric homeomorphisms rather than diffeomorphisms; see [**460**] for some physical speculations.

Pursuing an analogue of Borel–Weil is at least as interesting. Recall that for G compact, we get an action of G on line bundles on the flag manifold $G_{\mathbb{C}}/B$, and this accounts for the special (i.e. finite-dimensional) representations of G. Manin [**402**] suggested that something similar happens to $\mathfrak{Vir}$, with now the moduli spaces of curves playing the role of the flag manifold. This thought was made much more precise in [**357**], [**49**], [**13**]. Consider the enhanced moduli space $\widehat{\mathfrak{M}}_{g,n}$ of Section 2.1.4, where each of the n marked points on the genus-g surface is given a local coordinate z_i. A copy of $\mathfrak{Witt}$ for each marked point acts naturally on $\widehat{\mathfrak{M}}_{g,n}$: the vector field $z_i^{\ell}\partial/\partial z_i$, for $\ell \geq 1$, changes the coordinate z_i; $\partial/\partial z_i$ moves the ith point; and finally $z_i^{\ell}\partial/\partial z_i$ for $\ell \leq -1$ can change the conformal structure of the surface. This action fills out the tangent space to any point on $\widehat{\mathfrak{M}}_{g,n}$, i.e. we get a surjective Lie algebra homomorphism from $\mathfrak{Witt}$ to the tangent space at any point on $\widehat{\mathfrak{M}}_{g,n}$, and from this we can derive the central extension geometrically by considering determinant line bundles (a nice introduction to this important object is [**192**]) over $\widehat{\mathfrak{M}}_{g,n}$.

Pushing this much further would force us into the complexities (and riches) of algebraic geometry and $\mathcal{D}$-modules (see [**116**] for a gentle introduction to the simplest $\mathcal{D}$-modules). A far-reaching generalisation of the Borel–Weil Theorem is the equivalence of categories established by Beilinson–Bernstein and Brylinski–Kashiwara: given a Lie group G with semi-simple Lie algebra $\mathfrak{g}$, their 'localisation functor' relates an algebraic category, whose objects include the Verma modules of $\mathfrak{g}$, with a topological category of $\mathcal{D}$-modules (i.e. sheaves of modules over a ring of differential operators over the flag manifold $G_{\mathbb{C}}/B$). Describing this would take us far afield (see [**80**], [**417**] for reviews and references). In conformal field theory, the Virasoro algebra, moduli spaces $\mathfrak{M}_{g,n}$, and mapping class groups $\Gamma_{g,n}$ take the place of $\mathfrak{g}$, $G_{\mathbb{C}}/B$ and the Weyl group [**402**], [**530**]. [**49**] relates Virasoro modules to $\mathcal{D}$-modules on the enhanced moduli space $\widehat{\mathfrak{M}}_{g,n}$.

In any case, this deep relation between moduli spaces of curves and $\mathfrak{Vir}$ is significant to Moonshine, because of its relation to the analogues of the *Knizhnik–Zamolodchikov (KZ) equations* in any conformal field theory at any genus. We elaborate on this elsewhere (starting in Section 3.2.4), but for now let us say that 'chiral blocks' are sections over the moduli spaces $\widehat{\mathfrak{M}}_{g,n}$, and satisfy a system of partial differential equations saying roughly that they respect this $\mathfrak{Vir}$ action. The monodromy of those equations gives rise to projective actions of the mapping class groups on the spaces of chiral blocks. Now, the chiral blocks of the space $\mathfrak{M}_{1,1}$ (or rather $\widehat{\mathfrak{M}}_{1,1}$) are vertex operator algebra characters (including for instance (3.1.10)), and $\Gamma_{1,1} \cong \mathrm{SL}_2(\mathbb{Z})$ (or rather its central extension $\widehat{\Gamma}_{1,1} \cong \mathcal{B}_3$) acts on them. This is conformal field theory's explanation for the modularity of these characters. Thus the Virasoro algebra, through its action on the $\widehat{\mathfrak{M}}_{g,n}$, lies at the heart of Moonshine.

Question 3.1.1. (a) Let $G = \mathbb{Z}_2 \times \mathbb{Z}_2$. Define a map $P : G \to \mathrm{GL}_2(\mathbb{C})$ by

$$P(0,0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad P(1,0) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$
$$P(0,1) = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}, \quad P(1,1) = \begin{pmatrix} 0 & -\mathrm{i} \\ -\mathrm{i} & 0 \end{pmatrix}$$

Verify that P is a projective representation of G.
(b) Let Q be the order 8 'quaternion group', given by the following relations:

$$Q = \{\pm 1, \pm \mathrm{i}, \pm \mathrm{j}, \pm \mathrm{k} \mid -1 = (\pm \mathrm{i})^2 = (\pm \mathrm{j})^2 = (\pm \mathrm{k})^2,\ \mathrm{ij} = \mathrm{k} = -\mathrm{ji},\ -1 \text{ is in centre}\}.$$

Show that there is a homomorphism $\varphi : Q \to G$ with kernel $\{\pm 1\}$.
(c) Show that there is a true representation R of Q such that

$$P(x) = \delta(x)\, R(r(x)), \qquad \forall x \in G,$$

where $r(x) \in \varphi^{-1}(x)$, and where $\delta : G \to \mathbb{C}^\times$.

Question 3.1.2. Identify $G = S^1$ with $\mathbb{R}/\mathbb{Z}$, and for any class $[x] \in \mathbb{R}/\mathbb{Z}$, choose the unique representative $0 \le x < 1$. Verify that for any complex number α, the map $[x] \mapsto \alpha^x$ defines a one-dimensional projective representation of S^1. Find the corresponding true representation on the universal cover $\widetilde{G}$ of S^1.

Question 3.1.3. For this question, let G be the additive group $\mathbb{C}^2$. Define the function $\alpha : G \times G \to \mathbb{C}^\times$ by $\alpha(z, w) = \exp[z_2 w_1 - z_1 w_2]$. Verify that α obeys the 2-cocycle condition (3.1.1b), and construct the corresponding central extension.

Question 3.1.4. Find all one-dimensional central extensions of the Lie algebra A_1.

Question 3.1.5. Show that there are only two one-dimensional central extensions of the Witt algebra, up to isomorphism. (*Hint*: first show, changing basis if necessary, that $[L_0, L_n] = -nL_n$. Then consider anti-associativity of $[L_0[L_m L_n]]$.)

Question 3.1.6. (a) The group $\mathrm{PSL}_2(\mathbb{R})$ acts naturally on the unit disc $|z| < 1$ by Möbius transformations. Use this to embed $\mathrm{PSL}_2(\mathbb{R})$ naturally in $\mathrm{Diff}^+(S^1)$, and find the corresponding Lie subalgebra of $\mathrm{Vect}(S^1)$.
(b) The group $\mathrm{SL}_2(\mathbb{R})$ naturally acts on the space of semi-infinite rays $\mathbb{R}_{\ge}(x, y)$ in $\mathbb{R}^2$ with endpoint at the origin $(0, 0)$. Find this action, and use it to embed $\mathrm{SL}_2(\mathbb{R})$ in $\mathrm{Diff}^+(S^1)$. Find the corresponding Lie subalgebra of $\mathrm{Vect}(S^1)$.

Question 3.1.7. Prove that the Lie algebra of derivations of the algebra $\mathbb{C}[x^{\pm 1}]$ of Laurent polynomials is $\mathrm{Vect}(S^1)$.

Question 3.1.8. Find the constant $\alpha \in \mathbb{C}$ for which the new basis $L'_n = L_n + \alpha\,\delta_{n,0}C$ of $\mathfrak{Vir}$ has especially simple brackets $[L'_m, L'_n]$.

3.2 Affine algebras and their representations

The theory of nontwisted affine Kac–Moody algebras (usually called *affine algebras*) is very analogous to that of the finite-dimensional simple Lie algebras. Nothing infinite-dimensional tries harder to be finite-dimensional than affine algebras. Their construction is so trivial that it seems surprising anything interesting and new can happen here. But a certain 'miracle' happens. . .

Standard references for the theory of affine algebras are [**328**], [**337**], [**214**], [**551**]. We will ignore here an interesting part of the story: the KP hierarchy [**423**].

3.2.1 Motivation

Generalisations are too easy; they should be justified before they are endured. Here we describe the original justifications for the study of Kac–Moody algebras.

Each simple finite-dimensional Lie algebra has, as we know, a Weyl group, which is a symmetry of most of the data of the algebra (e.g. the weight multiplicities of finite-dimensional modules) and which encodes much (but not all) of the structure of the algebra. These Weyl groups are a very special sort of group: they are generated by reflections (namely those through the simple roots).

Associated with any vector $\alpha \in \mathbb{R}^n$, the reflection r_α through α, sending α to $-\alpha$ and fixing the hyperplane perpendicular to α is given by (1.5.5c). More abstractly, a reflection r is simply an *involution* (i.e. order 2: $r^2 = e$). A *finite reflection group* is a finite group generated by reflections. Coxeter studied these as symmetries of a regular solids.

For example, the dihedral group $\mathcal{D}_n$ (the group of symmetries of a regular n-gon) is a finite reflection group, consisting of n reflections and n rotations, and is generated by any two neighbouring reflections. The symmetric group $\mathcal{S}_n$ is a finite reflection group: it acts on an orthonormal basis e_i of $\mathbb{R}^n$ by permuting the subscripts, and is generated by the transpositions $(i, i+1)$, which are reflections r_{α_i} through the vector $e_i - e_{i+1}$.

Finite reflection groups have remarkably simple presentations.

Definition 3.2.1 *A* Coxeter group *G is a group with a set R of generators, whose complete list of relations is*

$$(rr')^{m(r,r')} = e, \qquad \forall r, r' \in R,$$

where $m(r,r) = 1$ and the other $m(r,r')$ all lie in $\{2, 3, \ldots, \infty\}$. (The value $m(r,r') = \infty$ means that rr' has infinite order.)

The geometry of Coxeter groups is quite pretty – see, for example, [**301**], [**84**]. In Section 7.1.1 we describe a generalisation due to Conway, and its relation to the Monster $\mathbb{M}$.

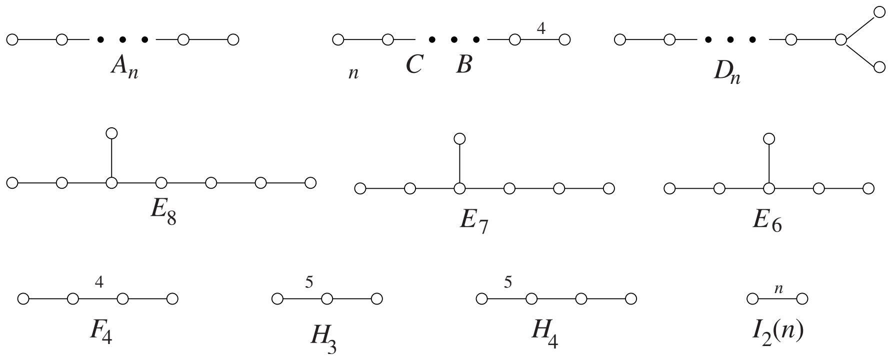

Fig. 3.1 The indecomposable finite Coxeter groups.

The list of finite Coxeter groups and finite reflection groups coincide. They are most easily described by the associated *Coxeter graph*: put a node for each generator $r \in R$, and connect two nodes with an edge labelled $m(r, r')$. To increase readability, erase the edge and label if $m(r, r') = 2$, and erase the label (but keep the edge) if $m(r, r') = 3$. The complete list of finite Coxeter groups (Coxeter, 1935) is given by arbitrary disjoint unions of the graphs of Figure 3.1. The group given by A_n is the symmetric group $\mathcal{S}_{n+1}$, and $I_2(n)$ is the dihedral group $\mathcal{D}_n$. The group H_3 is the symmetry group of the icosahedron, and is isomorphic to $\mathbb{Z}_2 \times \mathcal{A}_5$.

Figure 3.1 should remind us of Figure 1.17. Indeed, Figure 3.1 includes the Weyl groups of all simple finite-dimensional Lie algebras. More precisely, the Weyl groups consist of all finite Coxeter groups that obey the *crystallographic* condition: for all distinct $r, r' \in R$, $m(r, r') \in \{1, 2, 3, 4, 6\}$. Geometrically, the crystallographic condition says that the Coxeter group stabilises a lattice in $\mathbb{R}^n$ (see also Question 1.7.6). As we recall, the Weyl groups stabilise the corresponding root lattice.

Most Coxeter groups are infinite. As a graduate student, Robert Moody asked that, since the finite-dimensional semi-simple Lie algebras correspond to *finite crystallographic* Coxeter groups, what is the class of Lie algebras that correspond more generally to *any* Coxeter group? Presumably they should have a theory very similar to that of the semi-simple ones. The *partial* answer to Moody's beautiful question is that the Lie algebras corresponding to the (possibly infinite) *crystallographic* Coxeter groups are the Kac–Moody algebras! In fact, much of the interest in the affine algebras is due ultimately to their Weyl groups. We still don't know the Lie algebras corresponding to the noncrystallographic groups.

Victor Kac's road to these algebras was quite different. Let $\mathfrak{g}$ be a complex Lie algebra. By a $\mathbb{Z}$-*grading* we mean that we can write the vector space $\mathfrak{g}$ as $\mathfrak{g} = \oplus_{n=-\infty}^{\infty} \mathfrak{g}_n$, such that $[\mathfrak{g}_m, \mathfrak{g}_n] \subseteq \mathfrak{g}_{m+n}$ for all $m, n \in \mathbb{Z}$. We call $\mathfrak{g}$ a *simple* $\mathbb{Z}$-*graded* Lie algebra if, in addition, $\mathfrak{g}$ does not contain any nontrivial $\mathbb{Z}$-graded ideal.

It is probably hopeless to classify all simple $\mathbb{Z}$-graded Lie algebras – there are too many of them. However, decades earlier, Cartan had studied vector fields (i.e. derivations) on polynomial algebras, and found four infinite families that were simple $\mathbb{Z}$-graded, with

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

the dimension $\dim(\mathfrak{g}_n)$ bounded above by some polynomial in n. We say that these $\mathbb{Z}$-graded algebras have *polynomial growth*. Kac conjectured, and Olivier Mathieu proved, the complete list of such algebras.

Theorem 3.2.2 [409] *The simple $\mathbb{Z}$-graded Lie algebras of polynomial growth are:*
(a) *the finite-dimensional simple Lie algebras;*
(b) *the loop algebras (possibly twisted);*
(c) *Cartan's four families; and*
(d) *the Witt algebra $\mathfrak{Witt}$.*

The proof is long and complicated. We've already met the finite-dimensional $\mathfrak{g}$ and the Witt algebra. Cartan's algebras are defined explicitly in, for example, [**409**]. The 'loop algebras' are constructed next subsection (there are six infinite families and seven exceptionals).

What we call the affine algebras – our main interest this chapter – are the central extensions of these loop algebras. Of course, such algebras cannot be simple because of their centres, and for this reason aren't in Mathieu's list. In any case, the affine algebras (together with the Virasoro algebra) answer a technical but natural algebraic question.

A couple of years after their mathematical introduction [**325**, **430**], the nontwisted affine algebras were discovered independently in string theory [**42**], under the name *current algebras*.

The Lie algebras (a)–(d) in Mathieu's list are truly extraordinary, especially regarding their representation theory. The simplest of Cartan's families are the Weyl algebras, which are the differential operators on the algebra $\mathbb{C}[x_1, \ldots, x_n]$ of polynomials, generated by multiplication operators $x_1, \ldots, x_n$ and partial derivatives $\partial/\partial x_1, \ldots, \partial/\partial x_n$. Their modules are the simplest $\mathcal{D}$-modules and have deep connections throughout mathematics and physics (see [**116**], [**80**] for an introduction).

3.2.2 Construction and structure

Let $\overline{\mathfrak{g}}$ be any simple finite-dimensional Lie algebra. The affine algebra $\mathfrak{g} = \overline{\mathfrak{g}}^{(1)}$ is essentially the *(polynomial) loop algebra* $\mathcal{L}_{poly}\overline{\mathfrak{g}} = \mathbb{C}[t^{\pm 1}] \otimes \overline{\mathfrak{g}}$, defined to be all possible 'Laurent polynomials' $\sum_{n\in\mathbb{Z}} a_n t^n$ where each $a_n \in \overline{\mathfrak{g}}$ and all but finitely many $a_n = 0$. Treat t here as a formal variable. The bracket in $\mathcal{L}_{poly}\overline{\mathfrak{g}}$ is the obvious one: e.g. $[at^n, bt^m] = [ab]t^{n+m}$. Geometrically, $\mathcal{L}_{poly}\overline{\mathfrak{g}}$ is the Lie algebra of polynomial maps $S^1 \to \overline{\mathfrak{g}}$ (to see this realisation, think of $t = e^{2\pi i\theta}$). This explains the name, and also suggests several generalisations (e.g. take any manifold in place of S^1). But the loop algebra is simplest and best understood of these geometric Lie algebras, and the only one we consider in any depth (but see Section 3.3). Note that $\mathcal{L}_{poly}\overline{\mathfrak{g}}$ is infinite-dimensional. Its Lie groups are the *loop groups*, consisting of all maps of S^1 to a Lie group for $\overline{\mathfrak{g}}$ (Section 3.2.6).

We saw S^1 before, in the discussion of the Witt algebra, so we may expect the Virasoro and affine algebras to be related. In fact, the Witt algebra acts on the affine algebras

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

as *derivations*. By definition, a derivation D is a linear map that obeys the product rule for derivatives: $D([xy]) = [(Dx)y] + [x(Dy)]$. The easiest examples are the 'inner derivations': $D = \mathrm{ad}(x)$. All derivations of $\overline{\mathfrak{g}}$ are inner, but the loop algebra $\mathcal{L}_{poly}\overline{\mathfrak{g}}$ has several non-inner ones. In particular, because $\mathcal{L}_{poly}\overline{\mathfrak{g}}$ consists of all (polynomial) maps $S^1 \to \overline{\mathfrak{g}}$, the vector fields $\mathrm{Vect}_{poly}(S^1)$, and hence the Witt algebra $\mathfrak{Witt}$, act on it. More precisely, using the realisation $\ell_j = -t^{j+1}\mathrm{d}/\mathrm{d}t$ of the basis vectors of (1.4.9), we get the action

$$\ell_j.(at^n) = -t^{j+1}\frac{\mathrm{d}}{\mathrm{d}t}(at^n) = -nat^{j+n}. \tag{3.2.1}$$

This relation between $\mathfrak{Witt}$ and $\mathcal{L}_{poly}\overline{\mathfrak{g}}$ plays an important role in the whole theory.

The loop algebra has a unique nontrivial one-dimensional central extension $\widehat{\mathcal{L}_{poly}\overline{\mathfrak{g}}} = \mathcal{L}_{poly}\overline{\mathfrak{g}} \oplus \mathbb{C}C$, defined by

$$[t^m x, t^n y] = t^{m+n}[x, y] + m\delta_{m,-n}\,\kappa(x|y)\,C \tag{3.2.2a}$$

for all $x, y \in \overline{\mathfrak{g}}$ and $m, n \in \mathbb{Z}$, where $\kappa(x|y)$ is the invariant bilinear form (Killing form) of $\overline{\mathfrak{g}}$. Thus $\widehat{\mathcal{L}_{poly}\overline{\mathfrak{g}}}$ has the same relation to $\mathcal{L}_{poly}\overline{\mathfrak{g}}$ that $\mathfrak{Vir}$ has to $\mathfrak{Witt}$. Incidentally, [**344**] relates the central extensions (3.2.2a) and (3.1.5) to logarithms of differential operators.

In addition, for a technical reason (namely, to make the simple roots linearly independent, so weight spaces can be finite-dimensional), a further noncentral one-dimensional extension is usually made. The result: by the affine algebra $\mathfrak{g} = \overline{\mathfrak{g}}^{(1)}$ we mean the extension of $\widehat{\mathcal{L}_{poly}\overline{\mathfrak{g}}}$ by the derivation $\ell_0 := t\frac{d}{dt}$. The Witt algebra also acts naturally on $\mathfrak{g}$ (Question 3.2.3). The superscript '(1)' denotes the fact that the loop algebra was twisted by an order-1 automorphism, in other words that it is nontwisted. It is called 'affine' because of its Weyl group, as we shall see.

For example, elements in ${A_1}^{(1)}$ are triples $(a(t), w, x)$ where $w, x \in \mathbb{C}$ and $a(t) = \sum_{n\in\mathbb{Z}} a_n t^n$, for all $a_n \in \mathfrak{sl}_2(\mathbb{C})$ and only finitely many $a_n \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. The Lie bracket is

$$[(a(t), w, x), (a'(t), w', x')] = \left(\sum_{m,n} t^{m+n}[a_m, a'_n] + x\sum_n na'_n t^n - x'\sum_m ma_m t^m, \sum_m m\,\mathrm{tr}(a_m a'_{-m}), 0\right). \tag{3.2.2b}$$

Each object associated with $\overline{\mathfrak{g}}$ has an analogue here: Coxeter–Dynkin diagram, Weyl group, weights, ... For instance, the affine Coxeter–Dynkin diagram (Figure 3.2) is obtained from that of $\overline{\mathfrak{g}}$ (Figure 1.17) by adding one node, labelled with an 'x'. We have included the labels a_i and (where different from a_i) colabels $a_i^\vee$, whose significance is given next subsection.

The Cartan subalgebra $\mathfrak{h}$ plays the same role here that it does in Chapter 1: decomposing modules into weight spaces. It can be chosen to be $\overline{\mathfrak{h}} \oplus \mathbb{C}C \oplus \mathbb{C}\ell_0$, where $\overline{\mathfrak{h}}$ is a Cartan subalgebra of the semi-simple algebra $\overline{\mathfrak{g}}$. In fact, $\mathfrak{g}$ has a triangular decomposition $\mathfrak{g} =$

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

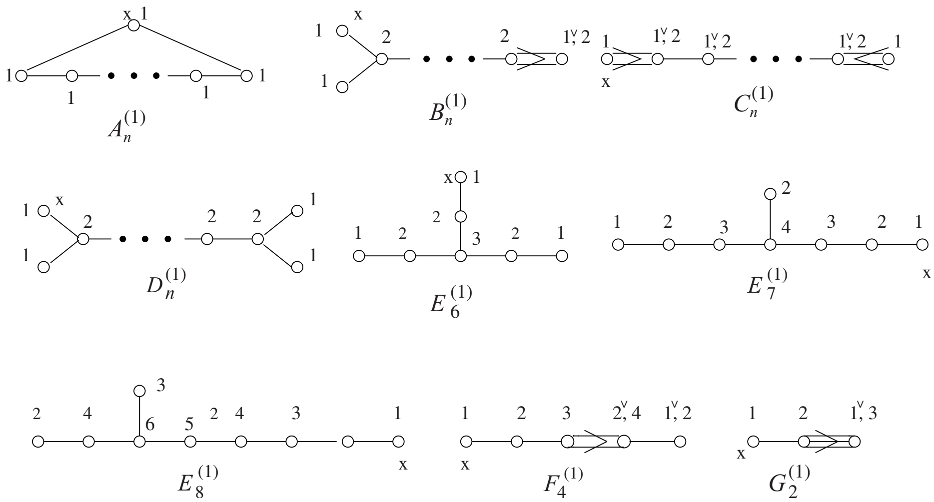

Fig. 3.2 The nontwisted affine Coxeter–Dynkin diagrams.

$\mathfrak{g}_+ \oplus \mathfrak{h} \oplus \mathfrak{g}_-$ (recall (1.5.5d)) where

$$\mathfrak{g}_\pm = \left(t^{\pm 1}\mathbb{C}[t^{\pm 1}] \otimes (\overline{\mathfrak{g}}_\mp \oplus \overline{\mathfrak{h}})\right) \oplus \mathbb{C}[t^{\pm 1}] \otimes \overline{\mathfrak{g}}_\pm \tag{3.2.3a}$$

and $\overline{\mathfrak{g}} = \overline{\mathfrak{g}}_+ \oplus \overline{\mathfrak{h}} \oplus \overline{\mathfrak{g}}_-$ is a triangular decomposition of $\overline{\mathfrak{g}}$. Given $\mathfrak{h}$, we obtain the root-space decomposition of $\mathfrak{g}$, as in (1.5.5a):

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{n\in\mathbb{Z}} \bigoplus_{\overline{\alpha}\in\overline{\Phi}} t^n \overline{\mathfrak{g}}_{\overline{\alpha}} \oplus \bigoplus_{n\in\mathbb{Z}\setminus 0} t^n \overline{\mathfrak{h}} \tag{3.2.3b}$$

where $\overline{\mathfrak{g}} = \overline{\mathfrak{h}} \oplus \oplus_{\overline{\alpha}} \overline{\mathfrak{g}}_{\overline{\alpha}}$. We return to (3.2.3) when we study $\mathfrak{g}$-modules next subsection, but for now note that if $\overline{\mathfrak{g}}$ has rank r, then the root spaces $t^n\overline{\mathfrak{h}}$ of $\mathfrak{g}$ have dimension r while all $t^n\overline{\mathfrak{g}}_{\overline{\alpha}}$ have dimension 1. The latter, which act like root spaces in $\overline{\mathfrak{g}}$, are called *real*, while the former are called *imaginary*.

This loop algebra construction can be twisted. Let $\overline{\mathfrak{g}}$ again be any simple and finite-dimensional Lie algebra and let $\mathfrak{g}$ be the corresponding affine algebra. Choose any symmetry α of the Coxeter–Dynkin diagram of $\overline{\mathfrak{g}}$, of order N say, and extend this into an automorphism of $\overline{\mathfrak{g}}$ as in Section 1.5.4. We can further extend α to an automorphism of $\mathfrak{g}$, by requiring α to fix C and ℓ_0, and send at^n to $\alpha(a)\xi_N^{-n}t^n$. Then the fixed-point subalgebra $\mathfrak{g}_0$ of $\mathfrak{g}$ is

$$\mathfrak{g}_0 = \left\{ \sum_n a_n t^n + wC + x\ell_0 \,\middle|\, a_n \in \overline{\mathfrak{g}}_{n \bmod N} \right\}, \tag{3.2.4}$$

where $\overline{\mathfrak{g}}_i$ are the eigenspaces of α in $\overline{\mathfrak{g}}$ (recall (1.5.12)). This Lie algebra $\mathfrak{g}_0$ is called a *twisted affine algebra* and is denoted $\overline{\mathfrak{g}}^{(N)}$. All twisted affine algebras are listed in Figure 3.3, with their colabels. Twisted affine algebras behave very analogously to the nontwisted ones, and also have a significant role in the theory (Section 3.4.1).

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

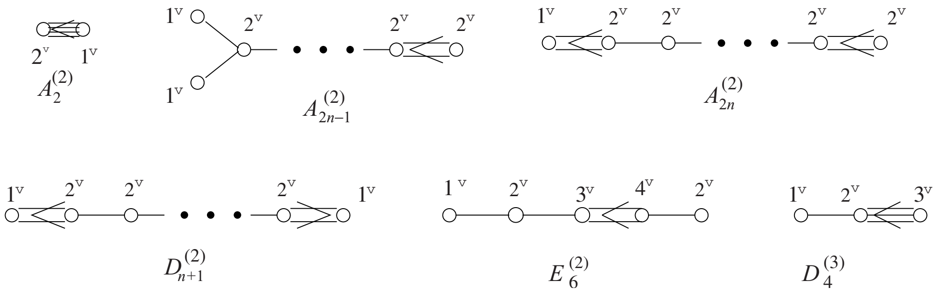

Fig. 3.3 The twisted affine Coxeter–Dynkin diagrams.

3.2.3 Representations

The loop algebra $\mathcal{L}_{poly}\overline{\mathfrak{g}}$ has no interesting modules, which is why we centrally extend it and introduce the affine algebras $\mathfrak{g} = \overline{\mathfrak{g}}^{(1)}$. No interesting $\mathfrak{g}$-module is finite-dimensional. However, $\mathfrak{g}$ has a triangular decomposition (3.2.3a), so highest-weight modules exist. Weights $\lambda \in \mathfrak{h}^*$ here are triples $(\overline{\lambda}, k, u) \in \overline{\mathfrak{h}}^* \times \mathbb{C}^2$; a weight vector v obeys

$$\overline{h}.v = \overline{\lambda}(h)\, v, \;\; C.v = kv, \;\; \ell_0.v = uv.$$

Define the Verma module $M(\overline{\lambda}, k, u)$ and the irreducible highest-weight module $L(\overline{\lambda}, k, u)$ – our greatest interest – as in Section 1.5.3. Given any highest-weight module M, the central term C acts as a multiple kI of the identity; this constant k is a fundamental invariant of the representation called the *level* of M. On the other hand, the value of u is irrelevant (at least when the level is not 0) – see Question 3.2.4.

A highest-weight module M is infinite-dimensional but comes with a grading $M = \oplus_{n=0}^{\infty} M_{u+n}$ into eigenspaces of ℓ_0. Because ℓ_0 commutes with $\overline{\mathfrak{g}}$, these spaces M_{u+n} are all $\overline{\mathfrak{g}}$-modules, and the lowest, namely M_u, has highest weight $\overline{\lambda}$. Using this we can define the graded-dimension as in (3.1.4b). However, the ℓ_0-spaces of Verma modules will be infinite-dimensional, as will those of $L(\overline{\lambda}, k, u)$ unless $\overline{\lambda} \in P_+(\overline{\mathfrak{g}})$. There are two ways to proceed: either find a more suitable grading, or (more important) consider instead the *character*.

Defining these characters requires decomposing our modules into weight-spaces, and for this we should fix a basis for $\mathfrak{h}^*$. A basis for $\mathfrak{h}$ is $h_1, \ldots, h_r$ (the usual basis for $\overline{\mathfrak{h}}$) together with $h_0 := C - \sum_{i=1}^{r} a_i^\vee h_i$ and $-\ell_0$ ($a_i^\vee$ are the colabels of Figure 3.2). The reason for introducing h_0 will be clearer in Section 3.3.1. The dual basis for $\mathfrak{h}^*$, corresponding to $h_0, \ldots, h_r, -\ell_0$, is written $\omega_0, \ldots, \omega_r, \delta$. Recall from Sections 1.4.3 and 1.5.2 the Killing form $\kappa(\overline{h}|\overline{h}')$ and $(\overline{\lambda}|\overline{\mu})$ for $\overline{\mathfrak{g}}$; its analogue for affine algebras (Question 3.2.5) obeys

$$\kappa(\overline{z} + a\ell_0 + uC|\overline{z}' + a'\ell_0 + u'C) = \kappa(\overline{z}|\overline{z}') - au' - ua', \tag{3.2.5a}$$

$$\left(\sum_{i=0}^{r} \lambda_i\omega_i + b\delta \,|\, \sum_{j=0}^{r} \mu_j\omega_j + d\delta\right) = \left(\sum_{i=1}^{r} \lambda_i\omega_i \,|\, \sum_{j=1}^{r} \mu_j\omega_j\right) + \sum_{i=0}^{r}(d\lambda_i + b\mu_i)\,. \tag{3.2.5b}$$

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

The level k is recovered from the weight λ by the formula

$$k = (\delta|\lambda) = \sum_{i=0}^{r} a_i^\vee \lambda_i. \tag{3.2.5c}$$

A useful formula gives the evaluation $\lambda(h)$:

$$\left(\sum_{i=0}^{r} \lambda_i \omega_i + b\delta\right)(\overline{z} + \tau \ell_0 + uC) = \left(\sum_{i=1}^{r} \lambda_i \omega_i\right)(\overline{z}) + ku - \tau b. \tag{3.2.5d}$$

In this notation, the roots of $\mathfrak{g}$ are $\overline{\alpha} - (\theta|\overline{\alpha})\omega_0 + n\delta$ for any root $\overline{\alpha}$ of $\overline{\mathfrak{g}}$ (these are the *real roots*, and have multiplicity 1), as well as $n\delta$ (the *imaginary roots*, with multiplicity equal to the rank r of $\overline{\mathfrak{g}}$). The root $\theta = \sum_{i=1}^{r} a_i \overline{\alpha}_i$ is called the *highest root* of $\overline{\mathfrak{g}}$, where a_i are the labels of $\mathfrak{g}$ (Figure 3.2). The positive roots are any of these with $n > 0$, together with $\overline{\alpha} - (\theta|\overline{\alpha})\omega_0$ for positive roots $\overline{\alpha}$. The simple roots are $\alpha_i := \overline{\alpha_i} - (\theta|\overline{\alpha}_i)\omega_0$ for $1 \le i \le r$, together with $\alpha_0 := \delta - \sum_{i=1}^{r} a_i \alpha_i$. Note that the adjoint representation of an affine algebra is not a highest-weight representation (why?). Many of these comments will make more sense when we associate a Coxeter–Dynkin diagram to $\mathfrak{g}$ in Section 3.3.1.

The weight-spaces for the Verma modules, and hence any highest-weight module M, are always finite-dimensional and so we can define their character ch_M as in (1.5.9a). For an easy example, the Verma module $M(\overline{\lambda}, k, 0) = M(\lambda)$ has character

$$\mathrm{ch}_{M(\lambda)}(h) = e^{\lambda(h)} \prod_{\alpha > 0} \left(1 - e^{-\alpha(h)}\right)^{-\mathrm{mult}\,\alpha}, \tag{3.2.6}$$

where '$\mathrm{mult}\,\alpha$' denotes the dimension of the root-space $\mathfrak{g}_\alpha$ (which now may be > 1). We can obtain convergent graded dimensions by specialising this in any number of ways; the most obvious (called the *principal gradation*) chooses $h \in \mathfrak{h}$ so that $e^{\alpha_i(h)} = x$ for all simple roots α_i ($0 \le i \le r$), and $e^{\omega_0(h)} = 1$ (x is a formal variable). In other words, the principal grading of a vector with weight $\lambda - \sum_{i=0}^{r} n_i \alpha_i$ is $\sum_i n_i$ less than the grading of λ – this gradation keeps track of how many 'creation operators' f_i (using notation introduced in Section 3.3.1) are applied to the 'vacuum' v in order to create the given state.

For example, the affine algebra $A_1{}^{(1)}$ has positive roots $2\omega_1 - 2\omega_0 + n\delta$ (for $n \ge 0$) as well as $m\delta$ and $-2\omega_1 + 2\omega_0 + m\delta$ (for $m > 0$). All root multiplicities are 1. The simple roots are $\alpha_1 = 2\omega_1 - 2\omega_0$ and $\alpha_0 = \delta - \alpha_1$. A highest-weight λ looks like $\lambda_0\omega_0 + \lambda_1\omega_1$, with level $\lambda_0 + \lambda_1$. Applying the principal gradation to the $A_1^{(1)}$-Verma module, its character (3.2.6) specialises to the principally-graded dimension

$$\begin{aligned}\dim^{pg}_{M(\lambda)}(x) &= x^{\lambda_1/2} \prod_{n=0}^{\infty} \left(1 - x^{-(1+2n)}\right)^{-1} \prod_{m=1}^{\infty} (1 - x^{-2m})^{-1} \prod_{m=1}^{\infty} \left(1 - x^{-(-1+2m)}\right)^{-1} \\ &= e^{-\pi \mathrm{i} \lambda_1 \tau} \eta(2\tau)/\eta(\tau)^{-2},\end{aligned} \tag{3.2.7}$$

where we write $x = e^{-2\pi \mathrm{i}\tau}$ and recall the Dedekind eta function from (2.2.6b). Thus once again we find the remarkable fact that graded dimensions of Verma modules have something to do with the modular group $\mathrm{SL}_2(\mathbb{Z})$ (compare (3.1.7)). Something similar happens for the highest-weight representations of any affine algebra!

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

Nothing particularly deep is happening here. The modularity of $\dim^{pg}$ arises here for free, simply from the combinatorics. Indeed, for any affine algebra, the specialised product of (3.2.6) is the generating function for some partition-like function as in (3.1.7b), and these have nice modular behaviour (by arguments like those used in Section 2.2.2).

More precisely, in the Verma module we get a free action of the creation operators of a Heisenberg subalgebra, coming from the central extension of the loop algebra of the Cartan subalgebra $\overline{\mathfrak{h}}$. Thus the modular group arises in affine algebra characters because of a Heisenberg algebra action. However, much as the discrete series (3.1.6) of $\mathfrak{Vir}$-modules behaves simpler than the other unitary $\mathfrak{Vir}$-modules, discretising the integral in (3.1.7d), certain families of $\mathfrak{g}$-modules have especially nice modular properties. What makes this work is the Weyl group. *It is this conjunction of the Heisenberg subalgebra with the affine Weyl group that makes affine algebras so special.*

The analogue for $\mathfrak{g}$ of the finite-dimensional modules of $\overline{\mathfrak{g}}$ are called the *integrable highest-weight modules*. Technically speaking, an integrable representation π is one where all $x \in \mathfrak{g}_\pm$ are locally nilpotent, that is, for each $v \in V$ there is a number $n_x(v)$ such that $\pi(x)^{n_x(v)}v = 0$. In particular, this means $e^{\pi(x)}$ is well-defined as an operator on the module by its Taylor series – in infinite dimensions most operators can't be exponentiated. These modules are called *integrable* because they are precisely those highest-weight modules that can be 'integrated' to a projective module of the corresponding loop group (Section 3.2.6). The integrable modules are precisely the unitary ones.

The *highest weight* $\lambda = \sum_{i=0}^{r} \lambda_i \omega_i$ is integrable iff each $\lambda_i \in \mathbb{N}$. Hence the set of all integrable level k highest weights is

$$P_+^k(\mathfrak{g}) := \left\{ \sum_{i=0}^{r} \lambda_i \omega_i \mid \lambda_i \in \mathbb{N},\ k = \sum_{i=0}^{r} a_i^\vee \lambda_i \right\}. \tag{3.2.8}$$

Simple formulae for the cardinality $\|P_+^k(\mathfrak{g})\|$ exist for all algebras (Question 3.2.6) – for example, for $A_r{}^{(1)}$ it is $\|P_+^k\| = \binom{k+r}{r}$. The most important weight in $P_+^k(\mathfrak{g})$ is $k\omega_0$, often denoted '0' in the literature. The module $L(k\omega_0)$ has a vertex operator algebra structure (Section 5.2.2) and corresponds to the vacuum sector in conformal field theory (Section 6.1.1).

The ℓ_0-eigenspaces of an integrable representation $L(\lambda)$ are all finite-dimensional representations of $\overline{\mathfrak{g}}$, and thus we can define its character $\mathrm{ch}_{L(\lambda)}$ as in (1.5.9a), although just as for the Virasoro algebra in (3.1.10) it proves to be more convenient to 'normalise' it:

$$\chi_\lambda(h) := e^{-(h_\lambda - c_\lambda/24)\,\delta(h)} \sum_{\beta \in \Omega(L(\lambda))} \dim L(\lambda)_\beta \, e^{\beta(h)}, \tag{3.2.9a}$$

where $L(\lambda) = \oplus L(\lambda)_\beta$ is the weight-space decomposition of $L(\lambda)$, $h \in \mathfrak{h}$, and

$$h_\lambda := \frac{(\lambda|\lambda + 2\rho)}{2\,(k + h^\vee)}, \tag{3.2.9b}$$

$$c_\lambda := \frac{k}{k + h^\vee} \dim \overline{\mathfrak{g}} \tag{3.2.9c}$$

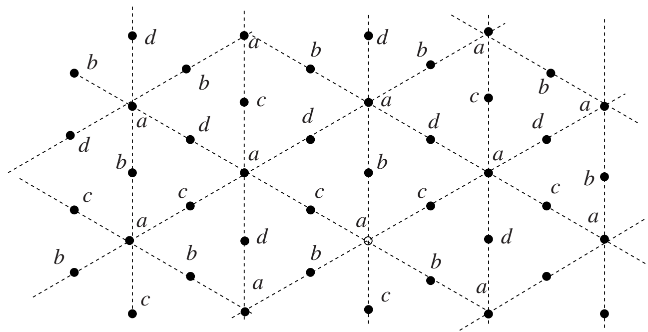

Fig. 3.4 The Weyl group of $A_2^{(1)}$ acting on level-2 weights.

are called the *conformal weight* and *central charge*, respectively, of $L(\lambda)$. The quantity $h^\vee = \sum_{i=0}^r a_i^\vee$ is called the *dual Coxeter number* and $\rho = \sum_{i=0}^r \omega_i$ the *Weyl vector*. The algebraic meaning of h_λ and c_λ involves the Virasoro algebra and is given shortly; $\delta(h)$ plucks out the coefficient $2\pi i\tau$ of ℓ_0 (recall (3.2.5d)). We are assuming in (3.2.9b) that the highest-weight component u has been set to 0 (Question 3.2.4). We discuss the normalisation (the exponential involving $h_\lambda - c_\lambda/24$) later in this subsection. As in (1.5.11), the character χ_λ can be written as an alternating sum over the Weyl group W, over a 'nice' denominator (namely the product in (3.2.6)). The difference is that W is now infinite.

See Figure 3.4 for the Weyl group of $A_2^{(1)}$ (projected to $\overline{\mathfrak{h}}^*$), and Question 3.2.7 for some simple calculations. Much of the interest in affine algebras can be traced to the 'miracle' that their Weyl groups are a semi-direct product $Q^\vee \times \overline{W}$ of translations in a lattice $\mathbb{Q}^\vee$ (the r-dimensional 'co-root lattice' of $\overline{\mathfrak{g}}$) with the (finite) Weyl group $\overline{W}$ of $\overline{\mathfrak{g}}$. More precisely, for any root $\overline{\alpha}$ of $\overline{\mathfrak{g}}$ define the *co-root* $\overline{\alpha}^\vee$ by $\overline{\alpha}^\vee = 2\overline{\alpha}/(\overline{\alpha}|\overline{\alpha})$; by the co-root lattice $Q^\vee \subset \overline{\mathfrak{h}}^* \subset \mathfrak{h}^*$ of $\overline{\mathfrak{g}}$ we mean the $\mathbb{Z}$-span of these co-roots. For any vector $\beta \in Q^\vee$, define the map

$$t_\beta(\mu) = \mu + (\mu|\delta)\,\beta - ((\mu|\beta) + (\beta|\beta)\,(\mu|\delta)/2)\,\delta, \tag{3.2.10a}$$

$\forall \mu \in \mathfrak{h}^*$. It is straightforward to verify $t_\beta t_\gamma = t_{\beta+\gamma}$, and thus these deserve the name 'translations'. Any element of the Weyl group W of $\mathfrak{g}$ can be written uniquely as a pair (t_β, w) for some $\beta \in Q^\vee$ and some $w \in \overline{W}$, and

$$(t_\beta, w) \circ (t_{\beta'}, w') = \left(t_\beta\, t_{w(\beta')}, ww'\right). \tag{3.2.10b}$$

As in (1.5.6d), weights $\mu \in \Omega(L(\lambda))$ in the same Weyl orbit of an integrable module have the same multiplicities. One thing this implies is that χ_λ will be of the form 'theta series'/denominator. In particular, the lattice is $Q^\vee$, and the '$(\beta|\beta)\delta$' term in (3.2.10a) provides the quadratic form in the lattice theta series. As we know from (2.3.10), theta series are modular forms, and this is the second complementary reason the modular group $SL_2(\mathbb{Z})$ makes an appearance (the first was the combinatorics of the free action of the Heisenberg subalgebra of creation operators). To make this more precise, consider

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

the highest weight $\lambda = \lambda_0\omega_0 + \lambda_1\omega_1 \in P_+^k(A_1^{(1)})$. Then

$$\chi_\lambda(2\pi \mathrm{i}\,(z+\tau\ell_0+uC)) = \frac{\Theta^{(k+2)}_{\lambda_1+1}(\tau,z,u) - \Theta^{(k+2)}_{-\lambda_1-1}(\tau,z,u)}{\Theta^{(2)}_{1}(\tau,z,u) - \Theta^{(2)}_{-1}(\tau,z,u)}, \quad (3.2.11a)$$

$$\Theta^{(n)}_m(\tau,z,u) := e^{-2\pi \mathrm{i} n u} \sum_{\ell\in\mathbb{Z}+\frac{m}{2n}} \exp[2\pi \mathrm{i} n\tau\ell^2 - 2\sqrt{2}\pi \mathrm{i} n\ell z]. \quad (3.2.11b)$$

Any $\mathfrak{g}$ has an analogue of (3.2.11), the *Weyl–Kac character formula*

$$\chi_\lambda(2\pi \mathrm{i}\,(z+\tau\ell_0+uC)) = \frac{\sum_{w\in\overline{W}} \epsilon(w)\,\Theta^{(k+h^\vee)}_{w(\lambda_1+\rho)}(\tau,z,u)}{\sum_{w\in\overline{W}} \epsilon(w)\,\Theta^{(h^\vee)}_{w(\rho)}(\tau,z,u)}, \quad (3.2.11c)$$

where both the numerator and denominator involve an alternating sum over the finite Weyl group $\overline{W}$ of $\overline{\mathfrak{g}}$, and where the theta series in (3.2.11c) involves a sum over the lattice $Q^\vee$ shifted by some weight and appropriately rescaled. For example, the Weyl group of A_1 is $\mathcal{S}_2$ and its co-root lattice $Q^\vee$ is $\sqrt{2}\mathbb{Z}$. The key variable in (3.2.11) is the modular one τ – the main role of the other variables is to ensure linear independence. The character χ_λ converges for any choice of $\tau \in \mathbb{H}$, $z \in \mathbb{C}^r$ and $u \in \mathbb{C}$.

Thus the denominator of the character of an irreducible integrable $\mathfrak{g}$-module $L(\lambda)$ is a modular form, by virtue of the combinatorics of Verma modules. The numerator is a modular form, by virtue of the structure and action of the affine Weyl group. Together they give a modular function.

Theorem 3.2.3 [**333**] *Let $\overline{\mathfrak{g}}$ be finite-dimensional and simple, and let $\mathfrak{g} = \overline{\mathfrak{g}}^{(1)}$ be the corresponding affine algebra. Define $\chi_\lambda(\tau,z,u) = \chi_\lambda(2\pi \mathrm{i}\,(z+\tau\ell_0+uC))$. Fix any level $k \in \mathbb{N}$. Then for any integrable weight $\lambda \in P_+^k(\mathfrak{g})$, $\chi_\lambda(\tau,0,0)$ is a modular function for some congruence subgroup $\Gamma(N)$. Moreover, define a column vector $\vec{\chi}(\tau,z,u)$ with entries $\chi_\lambda(\tau,z,u)$ for each $\lambda \in P_+^k(\mathfrak{g})$. Then there is a unitary representation ρ of $SL_2(\mathbb{Z})$ such that*

$$\vec{\chi}\left(\frac{a\tau+b}{f\tau+d}, \frac{z}{f\tau+d}, u - \frac{f\,(z|z)}{2(f\tau+d)}\right) = \rho\begin{pmatrix} a & b \\ f & d\end{pmatrix} \vec{\chi}(\tau,z,u),$$

for any $\begin{pmatrix} a & b \\ f & d\end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$.

We say that the characters χ_λ define a vector-valued Jacobi form for $\mathrm{SL}_2(\mathbb{Z})$, with multiplier ρ (recall Definition 2.2.2). This modularity of affine characters is fundamental to this book, and a prototypical example of much of what follows. The complex matrices $\rho(A)$ here are examples of modular data (Sections 6.1.2 and 6.2.1). A $\Gamma(N)$ that uniformly works in Theorem 3.2.3 is to let N be the least common multiple of all denominators of $h_\lambda - c/24$ (these will always be rational), as λ runs through the finite set $P_+^k(\mathfrak{g})$.

We can now explain McKay's observation (0.5.1) that the coefficients of $j(\tau)^{\frac{1}{3}}$ are related to the E_8 Lie group. $j^{\frac{1}{3}}(\tau)$ equals the character $\chi_{\omega_0}(\tau,0,0)$ of the integrable $E_8^{(1)}$-module. The q-coefficients ($q = e^{2\pi \mathrm{i}\tau}$) of $j^{\frac{1}{3}}(\tau)$ are thus dimensions of the

ℓ_0-eigenspaces of $L(\omega_0)$, which are automatically E_8-modules. Because $P_+^1(E_8^{(1)}) = \{\omega_0\}$, all modularity properties of the character $j^{\frac{1}{3}}$ are a direct consequence of Theorem 3.2.3.

All of this assumes the underlying finite-dimensional Lie algebra $\overline{\mathfrak{g}}$ is semi-simple. When it is merely reductive (i.e. the direct sum of copies of the one-dimensional abelian algebra $\mathfrak{u}_1$, with a number of simple Lie algebras), something different happens. For example, consider the affinisation of $\mathfrak{u}_1$ (the *oscillator algebra*). It has basis C, a_n ($n \in \mathbb{Z}$) and obeys relations

$$[C, a_n] = 0, \qquad [a_m, a_n] = m\delta_{m,-n}C. \tag{3.2.12a}$$

Its irreducible unitary modules are parametrised by a highest weight $\lambda \in \mathbb{R}$, and are Verma modules $M(\lambda)$. In particular, any $\lambda \in \mathbb{R}$ defines a different irreducible unitary module. They can be realised in the space of polynomials $\mathbb{C}[x_1, x_2, \ldots]$ by the operators $C.p(x) = p(x)$, $a_0.p(x) = \lambda\, p(x)$, and for all $n \geq 1$

$$a_n p(x) = \frac{\partial}{\partial x_n} p(x), \qquad a_{-n} p(x) = n x_n p(x). \tag{3.2.12b}$$

Note that the level k here is 1 (why can we demand $k = 1$?). The reader can verify that this representation has (normalised) character

$$\chi_\lambda(\tau) = q^{\lambda^2/2}/\eta(\tau). \tag{3.2.12c}$$

These characters aren't linearly independent (since $\chi_{-\lambda} = \chi_\lambda$), but the reader can work out the usual remedy. Their modularity is discussed in Section 6.2.2. In the language of conformal field theory, the unitary modules of the oscillator algebra ${\mathfrak{u}_1}^{(1)}$ are *quasi-rational* while the integrable modules of affine algebras are *rational*. Nevertheless, the oscillator algebra (studied in detail in [**334**]) is a convenient toy model for the affine algebras.

Last subsection we saw that $\mathfrak{Witt}$ acts naturally on loop algebras by derivations. Does $\mathfrak{Witt}$ act on affine modules? Consider the oscillator algebra for simplicity. We will have a universal $\mathfrak{Witt}$ action on ${\mathfrak{u}_1}^{(1)}$-modules M if we can construct the basis ℓ_n of (1.4.9) out of the operators a_m of (3.2.12a), that is realise the ℓ_n in the universal enveloping algebra $U({\mathfrak{u}_1}^{(1)})$ (or some completion thereof). We are led to consider quadratic combinations in the a_m, since that is the simplest after linear ones (which won't work), and also since ℓ_0 has the interpretation of a Hamiltonian, which always contains a quadratic part. Define

$$t_m = \sum_{i \in \mathbb{Z}} a_{-i} a_{m+i}. \tag{3.2.13a}$$

Being an infinite sum, convergence won't be automatic, but let's ignore that for now. Then

$$[t_m, a_n] = \sum_{i \in \mathbb{Z}} a_{-i}[a_{m+i}, a_n] + [a_{-i}, a_n] a_{m+i} = -n a_{m+n} C - n C a_{m+n} = -2n a_{m+n} C. \tag{3.2.13b}$$

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

In, for example, a highest-weight module, C acts as a scalar k, and so (at least for $k \neq 0$) $\ell_m := \frac{1}{2k} t_m$ mimics the action of the standard $\mathfrak{Witt}$ action $\ell_m = -t^{m+1}\mathrm{d}/\mathrm{d}t$ on the loop algebra $\mathcal{L}_{poly}\mathfrak{u}_1$. This looks promising. We compute from (3.2.13b)

$$\begin{aligned}[\ell_m, \ell_n] &= (2k)^{-2} \sum_{i\in\mathbb{Z}} [t_m, a_{-i}] a_{n+i} + (2k)^{-2} \sum_{j\in\mathbb{Z}} a_{-j} [t_m, a_{n-j}] \\ &= (2k)^{-1} \sum_{i\in\mathbb{Z}} i a_{m-i} a_{n+i} + (2k)^{-1} \sum_{j\in\mathbb{Z}} (-n-j) a_{-j} a_{m+n+j} = (m-n)\ell_{m+n},\end{aligned}$$

establishing that indeed the ℓ_m form a realisation of $\mathfrak{Witt}$ in $U(\mathfrak{u}_1{}^{(1)})$.

Unfortunately, the sum in (3.2.13a) doesn't converge. Take M to have highest-weight vector v with highest weight (λ, k). Then

$$t_0.v = \sum_{i\le -1} (a_i a_{-i} - iC).v + a_0 a_0.v + \sum_{j\ge 1} a_{-j} a_j.v = k^2 v + k \left(\sum_{j\ge 1} j \right) v, \quad (3.2.13c)$$

which diverges. This means (3.2.13a) must be modified. The simplest correction can be written $T_m := \sum_{i\in\mathbb{Z}} : a_{-i} a_{m+i} :$, where the *normal-ordering* $: a_m a_n :$ is defined to equal either $a_m a_n$ or $a_m a_n$, depending on whether or not $m \le n$. For $m \neq 0$, $T_m = t_m$, but $T_0.v = k^2 v$. Indeed, each operator T_m will be defined on any Fock space. We find that

$$L_m := (2k)^{-1} \sum_{i=-\infty}^{\infty} : a_{-i} a_{m+i} : \quad (3.2.14a)$$

satisfies both

$$[L_m, a_n] = -n a_{m+n}, \quad (3.2.14b)$$

$$[L_m, L_n] = (m-n) L_{m+n} + \frac{m^3 - m}{12} \delta_{m,-n}. \quad (3.2.14c)$$

Thus any highest-weight $\mathfrak{u}_1{}^{(1)}$-module is simultaneously a $\mathfrak{Vir}$-module with central charge $c = 1$. Thus this nonzero central charge arises as an analytic effect.

Using (3.2.12a), this normal-ordering (3.2.14) doesn't change $L_n = \ell_n$, for $n \neq 0$, but shifts the divergent ℓ_0 by the infinite multiple $\left(\sum_{i=1}^{\infty} i\right)$ of C. There is nothing particularly special about this normal-ordering; for example, for any fixed ℓ we could have replaced the condition '$m \le n$' with '$m \le n + \ell$', and nothing would have changed except L_0 would have been shifted by some other multiple of C. This is a clue to understanding what is so special about the $-c/24$ shifts in, for example, (3.1.10) or (3.2.9a). The arbitrariness of the normal-ordering can be removed by reinterpreting ('regularising') the divergent term in (3.2.13c) as $k\zeta(-1)$ (recall (2.3.1)). Equivalently, this amounts to replacing the normal-ordered L_0 with $L_0 - C/24$. This is the algebraic 'explanation' for the naturality of the shift, and hence the pervasive appearance of $-c/24$: simply put, algebra prefers $L_0 - C/24$ over all other combinations $L_0 + \alpha C$ (recall Question 3.1.8). It should thus not come as a complete surprise that so too does $\mathrm{SL}_2(\mathbb{Z})$. Incidentally, this '24', $\zeta(-1)$, the special dimensions $8 + 2$ and $24 + 2$ in string theory and the 24 of Section 2.5.1 are all directly related.

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

More generally, Bloch [**63**] considered other algebras of differential operators on S^1. In particular, in place of $\ell_n = -t^{n+1}\mathrm{d}/\mathrm{d}t$ he considers

$$\ell_n^{(r)} = (-1)^{r+1}(t\mathrm{d}/\mathrm{d}t)^r t^n (t\mathrm{d}/\mathrm{d}t)^{r+1}.$$

He obtains a (projective) realisation of these $\ell_n^{(r)}$ by normal-ordering operators in a Fock (or highest-weight) module, exactly as we do here. The analogue of $(m^3-m)/12$ in the bracket $[L_m^{(r)}, L_n^{(s)}]$ is a polynomial of degree $2r+2s+3$ in m. As before, we want to remove this arbitrary choice of normal-ordering. Naively dropping it introduces the divergence $1^{2r+1}+2^{2r+1}+\cdots$, so as before replace it with the Riemann zeta value $\zeta(-1-2r)$, i.e. replace $L_0^{(r)}$ with $L_0^{(r)} + (-1)^r\zeta(-1-2r)C/2$. Then the polynomial in m becomes the monomial $(r+s+1)!(r+s+1)!m^{2r+2s+3}/(2(2r+2s+3)!)$. This appearance of 'zeta function regularisation' in algebra has been interpreted and generalised in the vertex operator algebra framework (see [**375**] for a review).

Identical comments hold for affine algebras. Choose a basis x_a of $\overline{\mathfrak{g}}$, orthonormal with respect to the Killing form: $\kappa(x_a|x_b) = \delta_{ab}$. Then for $\lambda \in P_+^k(\mathfrak{g})$, (3.2.14) become

$$L_m := \frac{1}{2(k+h^\vee)}\sum_{j\in\mathbb{Z}}\sum_a :(t^{-j}x_a)(t^{m+j}x_a):, \tag{3.2.15a}$$

$$[L_m, xt^n] = -nxt^{m+n}, \qquad \forall x \in \overline{\mathfrak{g}}, \tag{3.2.15b}$$

$$[L_m, L_n] = (m-n)L_{m+n} + c_\lambda\frac{m^3-m}{12}\delta_{m,-n}. \tag{3.2.15c}$$

Thus the $\mathfrak{g}$-module $L(\lambda)$ is also automatically a completely reducible $\mathfrak{Vir}$-module. Each irreducible $\mathfrak{Vir}$-submodule has central charge c_λ and conformal weight $h \in h_\lambda + \mathbb{N}$ (see (3.2.9)). In $L(\lambda)$, the Virasoro generator L_0 and the derivation ℓ_0 of $\mathfrak{g}$ are related by $L_0 = h_\lambda Id + \ell_0$. Equation (3.2.15a), known as the *Sugawara construction*, should remind us of the *quadratic Casimir* $\Omega := \frac{1}{2}\sum_a x_a x_a$ of $\overline{\mathfrak{g}}$, that is, the simplest nontrivial element in the centre of $U(\overline{\mathfrak{g}})$; it acts on the irreducible $\overline{\mathfrak{g}}$-module $L(\overline{\lambda})$ as multiplication by the scalar $(\lambda|\lambda+2\rho)$ (recall (3.2.9b)). The shift by the dual Coxeter number $h^\vee$ in (3.2.15a) arises algebraically as the eigenvalue of Ω in the adjoint representation of $\overline{\mathfrak{g}}$; its physical significance is discussed in Section 6.2.1.

The integrable modules of twisted affine algebras ${X_r}^{(N)}$ (recall Figure 3.3) behave similarly. As we know from (3.2.4), ${X_r}^{(N)}$ is obtained from the nontwisted affine algebra $\mathfrak{g} = {X_r}^{(1)}$ and an order-N symmetry α of the Coxeter–Dynkin diagram of X_r. The integrable highest-weight $X_r^{(N)}$-modules $L(\lambda)$ are parametrised by $(r+1)$-tuples $\lambda \in P_+^k$ as in (3.2.8), where the co-labels $a_i^\vee$ are now given in Figure 3.3. These modules also have weight-space decompositions as in (1.5.6a) and characters χ_λ as in (3.2.9a). Their characters are also modular (see theorem 13.9 of [**328**] for details).

Theorem 3.2.4 [**333**] *The characters χ_λ, $\lambda \in P_+^k(A_{2r}^{(2)})$ form a vector-valued Jacobi function for $SL_2(\mathbb{Z})$, as in Theorem 3.2.3. For $\mathfrak{g} = {A_{2r-1}}^{(2)}, {D_{r+1}}^{(2)}, {E_6}^{(2)}$ and ${D_4}^{(3)}$, respectively, define $\mathfrak{g}' = {D_{r+1}}^{(2)}, {A_{2r-1}}^{(2)}, {E_6}^{(2)}, {D_4}^{(3)}$ and $N = 2, 2, 2, 3$; then the characters χ_λ, $\lambda \in P_+^k(\mathfrak{g})$, form a vector-valued Jacobi function for $\Gamma_0(N)$ (recall (2.2.4b)),*

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

and for each $\lambda \in P_+^k(\mathfrak{g})$,

$$\chi_\lambda\left(\frac{-1}{\tau}, \frac{z}{\tau}, u - \frac{(z|z)}{2\tau}\right) \in \mathrm{span}_{\mu' \in P_+^k(\mathfrak{g}')}\, \chi_{\mu'}\left(\frac{\tau}{N}, \frac{z}{N}, u\right).$$

3.2.4 Braided #3: braids and affine algebras

According to conformal field theory, the modularity of, for example, affine algebra characters arises through the monodromy of a system of partial differential equations (the Knizhnik–Zamolodchikov equations for a torus with one puncture). In this subsection we anticipate this important idea by considering the simpler and better-known situation of a sphere. See also [**355**], [**174**]; the basic idea of differential equation monodromy is nicely described in [**363**].

Theorem 3.2.5 *Consider a simply-connected open region* D *in* $\mathbb{C}$. *Consider the differential equation*

$$\frac{\mathrm{d}^2 w}{\mathrm{d}z^2} + P(z)\,\frac{\mathrm{d}w}{\mathrm{d}z} + Q(z)\,w = 0, \tag{3.2.16a}$$

where $P(z)$ *and* $Q(z)$ *are holomorphic in* D. *For any point* $z_0 \in D$, *and any* $\alpha, \beta \in \mathbb{C}$, *there is a unique function* $w(z)$, *holomorphic in* D, *satisfying the initial conditions*

$$w(z_0) = \alpha, \tag{3.2.16b}$$

$$\frac{\mathrm{d}w}{\mathrm{d}z}(z_0) = \beta. \tag{3.2.16c}$$

Hence the solutions w to (3.2.16a) form a two-dimensional space, parametrised by $\alpha, \beta \in \mathbb{C}$. For a proof of this theorem, see, for example, chapter XII of [**307**].

What if D is not simply-connected? One way to proceed would be to make D simply-connected by cutting it. For example, if D is $\mathbb{C}$ with n points $z_1, \ldots, z_n$ removed, then we can cut D along a non-self-intersecting polygonal path connecting $z_1, \ldots, z_n$ and ∞, avoiding the point z_0. Call D' the resulting simply-connected subregion of D. Then a holomorphic function on D restricts to a holomorphic function on D'; however, most holomorphic functions on D' won't extend continuously to D.

The other way to proceed is to consider the (simply-connected) universal cover $\pi : \widetilde{D} \to D$ (recall Section 2.1.2). We can then identify D with $\widetilde{D}/G$ for some group G isomorphic to the fundamental group $\pi_1(D)$; each $\gamma \in G$ is an automorphism of $\widetilde{D}$ shuffling the points $\widetilde{z} \in \pi^{-1}(z)$ above each $z \in D$. Functions h holomorphic on D lift to functions $h \circ \pi$ holomorphic on $\widetilde{D}$, although a typical function $\widetilde{h}$ on $\widetilde{D}$ won't correspond to a well-defined function on D. However, $\pi^{-1}(D') \subset \widetilde{D}$ consists of several connected open components, one for each $\gamma \in \pi_1(D)$, and through this there is a many-to-one correspondence between the holomorphic functions on D' and those on $\widetilde{D}$.

Let's return to the situation of Theorem 3.2.5, except with D now being non-simply-connected (although still connected). Then there is a unique solution w to (3.2.16) in D'. Writing $\widetilde{P} = P \circ \pi$ and $\widetilde{Q} = Q \circ \pi$, and choosing any $\widetilde{z}_0 \in \pi^{-1}(z_0)$, we can lift the equations (3.2.16) to $\widetilde{D}$ and again we obtain a unique solution $\widetilde{w}$, this time holomorphic

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

in $\widetilde{D}$. The space of solutions w on D', $\widetilde{w}$ on $\widetilde{D}$, are both two-dimensional. But we get more: both spaces carry naturally an action of the fundamental group $\pi_1(D)$, called the *monodromy representation*. More precisely, each automorphism $\gamma^* \in G \cong \pi_1(D)$ carries a solution $\widetilde{w}$ of (3.2.16a) to another solution $\widetilde{w} \circ \gamma^*$ – it preserves α, β but changes the choice $\widetilde{z_0} \in \pi^{-1}(z_0)$. It corresponds to an analytic continuation of w across the polygonal path cut out from D, along closed paths γ corresponding to γ^*.

A simple example should make this clear. Consider

$$\frac{\mathrm{d}^2 w}{\mathrm{d}z^2} + z^{-1}\frac{\mathrm{d}w}{\mathrm{d}z} = 0. \tag{3.2.17a}$$

Here, D is the punctured plane $\mathbb{C} \setminus \{0\}$ so we can take D' to be $\mathbb{C}$ with the negative real axis removed. The fundamental group $\pi_1(D)$ is $\mathbb{Z}$, and the universal cover $\widetilde{D}$ is the infinite spiral staircase. Two solutions to (3.2.17a) in D' are $w = \log z$ and $w = 1$. Analytically extend $w(z) = \log z$ along the unit circle starting at $z_0 = 1$ and running counterclockwise: as we cross the negative real axis continuity requires the value of w to be shifted by $2\pi\mathrm{i}$ from its previous 'principal' value. More generally, the path $\gamma^* = n$, winding n times around the origin, would pick up a monodromy of $2\pi\mathrm{i}n$. On the other hand, the constant solution $w(z) = 1$ is of course unchanged under analytic continuation. In terms of our basis $\{\log z, 1\}$, we thus obtain the monodromy representation

$$n \mapsto \begin{pmatrix} 1 & 2\pi\mathrm{i}\,n \\ 0 & 1 \end{pmatrix}. \tag{3.2.17b}$$

We are interested here in a slightly more complicated situation than that of Theorem 3.2.5. Let $\overline{\mathfrak{g}}$ be any finite-dimensional semi-simple Lie algebra and choose n distinct points $z_1, \ldots, z_n$ in $\mathbb{C}$. Recall the space $\mathfrak{C}_n$ defined in (1.2.6). Choose a basis x_a of $\overline{\mathfrak{g}}$, orthonormal with respect to the Killing form κ. For each i, choose a finite-dimensional $\overline{\mathfrak{g}}$-representation R_i, acting on a space V_i. Fix some complex number $\gamma \neq 0$. By the *Knizhnik–Zamolodchikov* (or *KZ*) equations we mean

$$\frac{\partial w}{\partial z_i} = \gamma \sum_{j \neq i} \sum_a \frac{R_i(x_a) \otimes R_j(x_a)}{z_i - z_j}\, w, \qquad 1 \leq i \leq n, \tag{3.2.18a}$$

where $w : \mathfrak{C}_n \to V_1 \otimes \cdots \otimes V_n$, and where $R_i(x_a)$, $R_j(x_a)$ act on the ith, jth components of the multilinear form w.

We recognise in (3.2.18a) the quadratic Casimir $\Omega = \sum_a x_a x_a$ discussed after (3.2.15). Physically (i.e. in the context of conformal field theory), w is a *chiral block* on the sphere $\mathbb{P}^1(\mathbb{C})$ with $n+1$ distinct marked points (namely $z_1, \ldots, z_n$ and $z_{n+1} = \infty$) for a Wess–Zumino–Witten model (Section 4.3.2). Geometrically (see e.g. [**338**]),

$$\frac{1}{2}\mathrm{d} - \gamma \sum_{\substack{j,i \\ i \neq j}} \sum_a R_i(x_a) \otimes R_j(x_a)\, \frac{\mathrm{d}z_i - \mathrm{d}z_j}{z_i - z_j} \tag{3.2.18b}$$

defines a *connection* (Section 1.2.2) on the trivial vector bundle $\mathfrak{C}_n \times W$, for $W = V_1 \otimes \cdots \otimes V_n$. An easy calculation verifies this connection is flat (i.e. has 0 curvature). The partial differential equations (3.2.18a) say that w is a *horizontal* or *parallel section*.

In other words, restricting to a simply-connected subregion $\mathfrak{C}_n'$ of $\mathfrak{C}_n$, the unique solution $w(z_1, \ldots, z_n)$ to (3.2.18a) satisfying some initial condition $w(z^{(0)}) = w^{(0)}$ is obtained geometrically by parallel-transporting the vector $w^{(0)}$ along any path γ in $\mathfrak{C}_n'$ connecting $z^{(0)}$ to the desired point $(z_1, \ldots, z_n)$.

Our context here is thus analogous to that of Theorem 3.2.5: parallel transport plays the role of analytic continuation, and the flatness of $\mathfrak{C}_n$ corresponds to the Monodromy Theorem of complex analysis (e.g. theorem 16.15 of [**481**]). The result is that the space of solutions to (3.2.18a) carries a representation of the fundamental group $\pi_1(\mathfrak{C}_n)$, i.e. of the pure braid group $\mathcal{P}_n$. We get an action of the full braid group through 'half-monodromies': a braid $\beta \in \mathcal{B}_n$ will take a solution w of (3.2.18a) to a solution of (3.2.18a) with values in $V_{\beta 1} \otimes \cdots \otimes V_{\beta n}$, where β acts on the indices $\{1, \ldots, n\}$ through the natural homomorphism $\phi : \mathcal{B}_n \to \mathcal{S}_n$ described in Section 1.1.4. In particular, if all V_i are isomorphic, the space of solutions of (3.2.18a) will carry a representation of the full group $\mathcal{B}_n$.

The infinitely many irreducible finite-dimensional modules of a simple Lie algebra naturally span a symmetric monoidal category (recall Section 1.6.2 for definitions); its character ring is isomorphic to a polynomial ring in r variables, where r is the rank of the algebra. On the other hand, the finitely-many level-k irreducible integrable modules of a nontwisted affine algebra span a braided monoidal category (in fact ribbon and modular categories); the corresponding character ring is called a *fusion ring* and is described in Section 6.2.1. The key ingredient in this category – the braiding – comes from this braid group monodromy. In Section 6.2.2 we see that this braid group monodromy, and associated braided monoidal category, generalise to the modules of sufficiently nice vertex operator algebras, and this (or if you prefer, conformal field theories) serves as the natural context for the modularity in Moonshine.

There are many other occurrences of the braid group in the mathematics and physics neighbouring Moonshine, and most of these are directly related to this KZ monodromy on a sphere. For example, the knot invariants arising from subfactors and quantum groups come from braid group representations, and Drinfel'd and Kohno have proved that these representations are the same ones coming from KZ monodromy.

On the other hand, the relation of the braid group $\mathcal{B}_3$ to $\mathrm{SL}_2(\mathbb{Z})$ and its modular functions, which we have seen already in Section 2.4.3 and which we argue later plays a fundamental role in Monstrous Moonshine, does not have a *direct* relation to this braid group monodromy. But we will see later that modularity too is due to monodromy of a system of partial differential equations – the analogue of these KZ equations for a once-punctured torus – defining a flat connection on the extended moduli space $\widehat{\mathfrak{M}}_{1,1}$. The solutions of these equations are spanned by the affine algebra characters (or more generally the vertex operator algebra one-point functions). The associated monodromy group is the mapping class group of $\widehat{\mathfrak{M}}_{1,1}$, which is readily seen to be $\mathcal{B}_3$.

Intriguingly, this means that we've come full circle. Poincaré's 125-year-old path to modular functions (see [**259**] for a review) was differential equations of the form (3.2.16a). Let $f(z)$, $g(z)$ be a basis for the space of solutions, and write $\xi(z) = f(z)/g(z)$. Note that the monodromy group acts on ξ by Möbius transformations: $\xi \mapsto \frac{a\xi+b}{c\xi+d}$.

Poincaré found that, at least in some cases, when we invert $\xi(z)$ and write z as a function of ξ, then z will be a modular function for some discrete subgroup of $SL_2(\mathbb{R})$, acting on $\xi \in \mathbb{H}$.

A simple example is Legendre's equation

$$\frac{d^2y}{dk^2} + \frac{1-3k^2}{k(1-k^2)}\frac{dy}{dk} - \frac{y}{1-k^2} = 0.$$

This has the elliptic periods $K(k)$ and $K'(k) = K(k')$ as solutions (recall Section 2.2.1). It is more convenient to change variables to $z = k^2$, when this equation becomes

$$\frac{d^2w}{dz^2} + \frac{1-2z}{z(1-z)}\frac{dw}{dz} - \frac{w}{4z(1-z)} = 0. \tag{3.2.19a}$$

Then $K'(z) = K(1-z)$, since $k^2 + k'^2 = 1$. The domain D is the plane with $z = 0$ and $z = 1$ removed; its fundamental group π_1 is the free group $\mathcal{F}_2 = \langle \sigma_0, \sigma_1 \rangle$ generated by counter-clockwise loops σ_k about $z = k$. It turns out that $K(z)$ is holomorphic at $z = 0$, but $K'(z)$ has a logarithmic singularity there: $K'(z) + \frac{1}{\pi}K(z)\log z$ is holomorphic at $z = 0$. Thus as we go counter-clockwise in a small circle about $z = 0$, $K(z)$ is unchanged but $K'(z)$ becomes $K'(z) - 2\mathrm{i}K(z)$. Hence, as we go counter-clockwise in a small circle about $z = 1$, $K'(z)$ is unchanged but $K(z)$ becomes $K(z) + 2\mathrm{i}K'(z)$. Thus in terms of the basis $\{K(z), \mathrm{i}K'(z)\}$ of solutions to (3.2.19a), the monodromy representation becomes

$$\sigma_0 \mapsto \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \qquad \sigma_1 \mapsto \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}. \tag{3.2.19b}$$

For the details of this calculation, see chapter 14.5 of [**486**]. The image of (3.2.19b) is precisely the congruence subgroup $\Gamma(2)$, which indeed is isomorphic to $\mathcal{F}_2$. Now, Poincaré would have us invert the function $\mathrm{i}K'(z)/K(z)$. That ratio turns out to always be in $\mathbb{H}$, and so denote it $\tau(z)$. Expressing $z = k^2$ as a function of τ, we obtain

$$z(\tau) = \frac{\theta_2(\tau)^4}{\theta_3(\tau)^4}. \tag{3.2.19c}$$

Indeed, we know from (2.3.8) that (3.2.19c) is invariant under $\Gamma(2)$.

It is remarkable to recover in this way the group $\Gamma(2)$, its action on $\mathbb{H}$ and a modular function for $\Gamma(2)$ (in fact, $\Gamma(2)$ is genus-0 and θ_2^4/θ_3^4 generates all of its modular functions). There are many other examples of this kind, for example

$$w'' + z^{-1}w' + \left(\frac{31}{144}z - \frac{1}{36}\right) z^{-2}(z-1)^{-2}w = 0$$

yields in this way the j-function. See [**516**] for more on the deep relation between modular forms and hypergeometric functions. The relations between affine algebras, the KZ equation and hypergeometric functions is explored in [**541**]. The Riemann–Hilbert problem asks that all linear representations of mapping class groups arise as monodromies; see the appendix of [**259**] for a history of this problem and chapter VIII of [**80**] for the modern treatment and generalisation using $\mathcal{D}$-modules.

Thus Poincaré, like conformal field theory over a century after him, finds it natural to interpret modularity using differential equation monodromy!

3.2.5 Singularities and Lie algebras

In this subsection we quickly review the geometry underlying the associations of singularities to simple Lie algebras (duVal) and affine Lie algebras (McKay), which are described in Section 2.5.2. This is related to mirror symmetry and provides a new explanation for the modularity of affine algebra characters.

Let Γ be a finite subgroup of $SU_2(\mathbb{C})$. Then the orbifold $\mathbb{C}^2/\Gamma$ has a critical point at the fixed point $(0, 0)$; the minimal resolution $X_\Gamma = \widetilde{\mathbb{C}^2/\Gamma}$ is a smooth noncompact real 4-manifold with an ALE ('asymptotically locally Euclidean') hyper-Kähler structure. An ALE manifold is Riemannian, with a metric tending quickly to the Euclidean one as $r \to \infty$. Physically, they correspond to positive-definite self-dual solutions to Einstein's gravitation equations in a vacuum ('gravitational instantons'). Conversely, any ALE hyper-Kähler manifold is diffeomorphic to some X_Γ for a unique Γ. The details are reviewed in [**362**].

Kronheimer–Nakajima [**362**] use the Atiyah–Singer Index Theorem to directly relate the duVal and McKay data associated with a simple singularity. Let X be an ALE hyper-Kähler manifold and $\Gamma < SU_2(\mathbb{C})$ the corresponding finite group. Then asymptotically at infinity, X is flat and in fact looks like $\mathbb{R}^4/\Gamma$. Given any vector bundle E over X, the fibre over ∞ defines a Γ-module R via monodromy. Kronheimer–Nakajima take E to be $\mathcal{R} \otimes \mathcal{R}^*$, where $\mathcal{R}$ is the tautological vector bundle, because its index vanishes. Then the monodromy representation R decomposes as $\sum_i \rho_i \otimes \overline{\rho_i}$, where ρ_i are the irreducible representations of Γ. The Index Theorem provides an expression for the numbers

$$\frac{1}{\|\Gamma\|} \sum_{\gamma \in \Gamma, \gamma \neq e} \frac{\mathrm{ch}_{\rho_i}(\gamma)\, \mathrm{ch}_{\rho_j^*}(\gamma)}{2 - \mathrm{ch}_\rho(\gamma)},$$

for $i, j = 0, 1, \ldots, n$, as an integral over X involving the intersection matrix, where ρ is the defining two-dimensional representation of $\Gamma < SL_2(\mathbb{C})$. From this they quickly establish the equivalence of duVal's observation that the intersection matrix is the negative of the $n \times n$ Cartan matrix, with McKay's interpretation of the $(n + 1) \times (n + 1)$ Cartan matrix as coefficients of the product $\rho \otimes \rho_i$.

The first direct relation between simple singularities and the Lie algebras A_r, D_r, E_r was established by Brieskorn [**86**]. Let $\overline{\mathfrak{g}}_\Gamma$ be the finite-dimensional simple Lie algebra associated with Γ, and G_Γ the corresponding Lie group. Let W be its (finite) Weyl group, and choose any Cartan subalgebra $\overline{\mathfrak{h}}$. Then Brieskorn obtained the singularity $\mathbb{C}^2/\Gamma$ and its resolution by studying the map $\overline{\mathfrak{g}}_\Gamma \to \overline{\mathfrak{h}}/W$, sending $x = x_s + x_n \in \overline{\mathfrak{g}}_\Gamma$ (this decomposition of x is just the Jordan canonical form [**300**]) to the orbit of the semi-simple part x_s under the adjoint action of G_Γ – these orbits are parametrised by $\overline{\mathfrak{h}}/W$ (Section 1.5.2).

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

More relevant for us is Nakajima's geometric realisation of affine algebras and their integrable representations (see e.g. his review [**445**]). Let E be an anti-self-dual Yang–Mills instanton over X with gauge group $\mathrm{U}_k(\mathbb{C})$. These bundles E are associated with three discrete invariants: the monodromy representation R as above; the first Chern class $c_1(E)$; and the instanton number $ch_2(E) \in \mathbb{N}$.

The monodromy R is a k-dimensional representation of Γ. Decompose R into irreducibles: $R = \sum_i \lambda_i \rho_i$, where the multiplicities $\lambda_i \in \mathbb{N}$. Then taking dimensions we obtain $k = \sum_{i=0}^n a_i \lambda_i$, where $a_i = \dim \rho_i$. According to the McKay correspondence, a_i are the labels of the corresponding nontwisted affine algebra $\mathfrak{g}_\Gamma$, and so $\lambda = \sum_i \lambda_i \omega_i$ is a level-k integrable highest weight of $\mathfrak{g}_\Gamma$.

Nakajima proceeds to construct not only $\mathfrak{g}_\Gamma$ from the geometric data, but also the $\mathfrak{g}_\Gamma$-module $L(\lambda)$. The singularity at $(0,0)$ of $\mathbb{C}^2/\Gamma$ resolves locally into n copies of the sphere $\mathbb{P}^1(\mathbb{C})$. These give a basis of $H_2(X,\mathbb{Z})$; Nakajima identifies them with the usual basis h_i of a Cartan subalgebra of the finite-dimensional algebra $\overline{\mathfrak{g}_\Gamma}$ and their intersection form with the Killing form. Thus the dual vectors $c_1(E)$ are weights. The number $ch_2(E)$ is identified with an eigenvalue of the derivation $\delta = L_0$. The other generators e_i, f_i of $\mathfrak{g}_\Gamma$ can be interpreted likewise. The moduli space $\mathcal{M}(k)$ of $\mathrm{U}_k(\mathbb{C})$-instantons on X has a finite-dimensional connected component $\mathcal{M}(k)_{\lambda,\mu,n}$ for every choice of monodromy λ, $c_1 = \mu$ and $ch_2 = n$. The infinite-dimensional cohomology space $H^\star(\mathcal{M}(k))$ carries a natural though reducible module of the affine algebra $\mathfrak{g}_\Gamma$. However, the middle-dimensional cohomology

$$\oplus_{\mu,n} H^d(\mathcal{M}(k)_{\lambda,\mu,n}), \qquad d = \frac{1}{2}\dim(\mathcal{M}(k)_{\lambda,\mu,n}) \tag{3.2.20}$$

is isomorphic to $L(\lambda)$, with each summand being a weight-space (the middle-dimensional cohomology spaces are generally the most interesting – for example, the pairing defines a bilinear form, here the Killing form, on them).

This construction generalises considerably [**445**]. It also has a natural interpretation in string theory. The *Bogomol'nyi–Prasad–Sommerfeld ('BPS') states* generally form an algebra closely related to Borcherds–Kac–Moody algebras (Section 3.3.2) [**276**]. Inside this BPS algebra for the heterotic string on the torus T^4 is the associated affine algebra. This string theory is dual to that of a type IIA string on a K3 surface (X_Γ is essentially a noncompact K3), where Nakajima's construction is very natural. So string theory interprets Nakajima's cohomological construction of affine algebras as a manifestation of mirror symmetry [**276**]. In this context, Vafa–Witten suggested that the modularity of affine algebra characters may have to do with S-duality [**540**], an $\mathrm{SL}_2(\mathbb{Z})$-symmetry of the heterotic string. It seems unlikely though that this can account for the modularity in arbitrary RCFT. We revisit mirror symmetry [**291**] in Section 7.3.8.

Physically, instantons are configurations for which the classical action (4.1.3) has a local minima. This means that in the corresponding quantum theory, we should perturb about them just as we do about the vacua. See the review [**159**] on instantons in supersymmetric theories. It turns out that (not necessarily holomorphic) modular forms appear naturally in this context, with the modular group arising again through S-duality.

Recalling Section 2.4.3, we can ask: *Can S-duality sometimes be extended naturally into a $\mathcal{B}_3$ symmetry?* This may provide a universal simplification, for example, for fractional instantons.

3.2.6 Loop groups

This brief subsection introduces the Lie groups of the affine algebras, by translating the previous subsections into this geometric language. See the book [**465**] for more details. Loop groups appear directly in Wess–Zumino–Witten string theory, and in the study of certain differential equations (solitons), but otherwise the affine algebra is mathematically prior. From our (limited) perspective, the geometric insight gained isn't obviously worth the analytic subtleties.

Choose any compact Lie group G, and let $\overline{\mathfrak{g}}$ be its Lie algebra. By the *loop group* $\mathcal{L}G$ we mean all smooth maps $S^1 \to G$, and by the *loop algebra* $\mathcal{L}\overline{\mathfrak{g}}$ we mean all smooth maps $S^1 \to \overline{\mathfrak{g}}$. The loop group $\mathcal{L}G$ has a group structure given by pointwise product, and in fact it forms an infinite-dimensional Lie group with Lie algebra $\mathcal{L}\overline{\mathfrak{g}}$.

Think of G as a subgroup of $U_n(\mathbb{C})$, as we can. The *polynomial loop group* $\mathcal{L}_{poly}G$ is the set of all loops $\gamma \in \mathcal{L}G$ that can be written in the form

$$\gamma(z) = \sum_{m=-\infty}^{\infty} a_m z^m,$$

i.e. as a matrix-valued function, where $z \in S^1$ and each a_m is an $n \times n$ complex matrix, with all but finitely many $a_m = 0$. Note that $\mathcal{L}_{poly}G$ is indeed a group – for example, inverse is given by $\gamma(z)^{-1} = \sum_m a_m^\dagger z^{-m} \in \mathcal{L}_{poly}G$. However, note that $\mathcal{L}_{poly}S^1$ consists of the monomials az^m for some constants $m \in \mathbb{Z}$ and $a \in S^1 \subset \mathbb{C}$ (to see this, multiply $\gamma(z)$ by $\gamma(z)^\dagger$; the result is a Laurent polynomial in z with coefficients in $\mathbb{C}$, which identically equals 1 for uncountably many $z \in \mathbb{C}$). Thus $\mathcal{L}_{poly}S^1$ has Lie algebra $\mathrm{i}\mathbb{R} \neq \mathcal{L}_{poly}S^1$. For semi-simple G, however, $\mathcal{L}_{poly}G$ has Lie algebra $\mathcal{L}_{poly}\overline{\mathfrak{g}}$, as we'd like.

The loop group $\mathcal{L}G$ is generally better behaved than $\mathcal{L}_{poly}G$. For example, we know the exponential map exp: $\overline{\mathfrak{g}} \to G$ is onto and locally one-to-one. The exponential map $\mathcal{L}\overline{\mathfrak{g}} \to \mathcal{L}G$ is defined in the obvious way (as the exponential of a matrix-valued function), and it is locally (but not globally) both one-to-one and onto. On the other hand, the exponential of a Laurent polynomial will usually not be a Laurent polynomial, and so the exponential map doesn't exist for polynomial loops. By way of comparison, as we mentioned in Section 3.1.2, $\exp : \mathrm{Vect}(S^1) \to \mathrm{Diff}(S^1)$ is neither locally one-to-one nor locally onto (in fact its image is nowhere dense).

$\mathrm{Diff}(S^1)$ acts naturally on $\mathcal{L}G$, by changing the parametrisation of the loop (for simple G, the only other automorphisms of $\mathcal{L}G$ come from the loop group of $\mathrm{Aut}(G)$).

To enrich the representation theory of $\mathcal{L}G$, we centrally extend $\mathcal{L}G$ by S^1. For simple G, $\mathcal{L}G$ has an inequivalent central extension $\widetilde{\mathcal{L}G}_n$ for each $n = 0, 1, 2, \ldots,$ and these exhaust all of them. $\widetilde{\mathcal{L}G}_0 \cong S^1 \times \mathcal{L}G$ is the trivial extension; $\widetilde{\mathcal{L}G}_1$ is the unique simply-connected such extension. $\widetilde{\mathcal{L}G}_n$ is obtained from $\widetilde{\mathcal{L}G}_1$ by quotienting by the order-n

subgroup of the centre S^1. The Lie algebra of any $\widetilde{\mathcal{L}G}_n$, $n > 0$, is isomorphic to the unique nontrivial central extension of the loop algebra $\mathcal{L}\overline{\mathfrak{g}}$.

We're interested in continuous projective representations of $\mathcal{L}G$ by bounded operators in a Hilbert space $\mathcal{H}$. We want these as usual to be $\mathbb{Z}$-graded. But an S^1 action is the same as a $\mathbb{Z}$-grading. More precisely, consider the group S^1 of rigid rotations R_θ in $\mathcal{L}G$ – that is, a loop $\gamma(t) \in \mathcal{L}G$ gets sent to the loop $(R_\theta\gamma)(t) = \gamma(t - \theta)$ for some fixed $0 \leq \theta < 2\pi$. We can decompose this S^1 action on $\mathcal{H}$ Fourier-like into (the completion of) a direct sum

$$\oplus_{\ell=-\infty}^{\infty}\mathcal{H}(\ell)$$

of subspaces $\mathcal{H}(\ell)$ on which R_θ acts like $e^{-\mathrm{i}\ell\theta}$. In other words, $e^{-\mathrm{i}L_0\theta}$ represents R_θ.

We require $\mathcal{H}(\ell)$ to vanish for all ℓ sufficiently close to $-\infty$. Because of the conformal field theory interpretation given next chapter, these eigenvalues ℓ are thought of as energy, and these representations are called *positive energy representations*. Any such projective representation of $\mathcal{L}G$ lifts to one of the semi-direct product of this S^1 with any central extension $\widetilde{\mathcal{L}G}_n$. This double S^1-extension of $\mathcal{L}G$ corresponds to the double $\mathbb{C}$-extension of the (polynomial) loop algebra performed in Section 3.2.2.

Let G be semi-simple. Any projective representation $\mathcal{H}$ of $\mathcal{L}G$ of positive energy is unitary and hence is completely reducible into a discrete direct sum of irreducible representations. The above action of S^1 (through the operators $e^{-\mathrm{i}L_0\theta}$) extends to a projective action of $\mathrm{Diff}^+(S^1)$. The L_0-eigenspaces $\mathcal{H}(\ell)$ of any irreducible representation $\mathcal{H}$ are all finite-dimensional. We can refine these eigenspaces by choosing a maximal torus T of G (it will be isomorphic to $S^1 \times \cdots \times S^1$ (r times), where r is the rank of G). We can diagonalise this action of $S^1 \times T \times S^1$, where the first S^1 is from the rigid rotations, and the second from the central extension; then

$$\mathcal{H}(n) = \oplus_{\mu\in P_+(\overline{\mathfrak{g}})}\mathcal{H}(n, \mu, k)$$

is the corresponding diagonalisation into weight spaces. Of course we are rediscovering the weight-space decomposition in, for example, (3.2.9a). The 'rigid rotation' S^1 corresponds to the extension of the loop algebra $\mathcal{L}_{poly}\overline{\mathfrak{g}}$ by the derivation $-\ell_0$, and the projective $\mathrm{Diff}^+(S^1)$ action corresponds to the Virasoro action (3.2.15). The maximal torus $S^1 \times T \times S^1$ of the double extension of $\mathcal{L}G$ corresponds to the (real) Cartan subalgebra $\mathfrak{h}$ of $\overline{\mathfrak{g}}^{(1)}$. Given any irreducible projective representation of $\mathcal{L}G$ of positive energy, then the derived projective representation of $\mathcal{L}\overline{\mathfrak{g}}$, restricted to $\mathcal{L}_{poly}\overline{\mathfrak{g}}$, is an integrable highest-weight representation $L(\lambda)$ of $\overline{\mathfrak{g}}^{(1)}$. Conversely, any such representation of $\overline{\mathfrak{g}}^{(1)}$ lifts ('integrates') to a projective representation of positive energy of $\mathcal{L}G$.

Any irreducible projective representation of $\mathcal{L}G$ lifts to a true representation of the simply-connected $\widetilde{\mathcal{L}G}_1$. It lifts to a true representation of $\widetilde{\mathcal{L}G}_n$ iff n divides the level k.

The analogue of Borel–Weil applies here much as in Section 1.5.5; the role of the symmetric space G/T is played here by the infinite Grassmannian $\mathcal{L}G/T$ (see chapter 11 of [**465**]). The irreducible representations also fit in well with Kirillov's orbit method [**198**].

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

It is tempting to hope more generally that the group Map(M, G), for any manifold M and compact G, should be a relatively accessible class of infinite-dimensional Lie groups. However, the theory is much more difficult than $\mathcal{L}G = \text{Map}(S^1, G)$ and little is known about their representations (see chapter 3 and section 9.1 in [**465**]).

Question 3.2.1. Define A to be the space of all differential operators of the form $\sum_{m,n\in\mathbb{Z}} a_{m,n} x^m \mathrm{d}^n/\mathrm{d}x^n$, where all $a_{m,n} \in \mathbb{C}$ and all but finitely many $a_{m,n}$ equal 0. Define a Lie algebra structure on A in the obvious way. Prove that A is a simple $\mathbb{Z}$-graded Lie algebra of polynomial growth.

Question 3.2.2. For a manifold X and Lie algebra L, when is Map(X, L) a simple Lie algebra?

Question 3.2.3. Show that the Witt algebra acts on the affine algebra $\overline{\mathfrak{g}}^{(1)}$ as derivations.

Question 3.2.4. Show that a highest-weight representation of a nontwisted affine Lie algebra $\mathfrak{g} = X_r{}^{(1)}$ with highest weight $(\overline{\lambda}, k, u)$ is isomorphic as a $\mathfrak{g}$-module to one with highest weight $(\overline{\lambda}, k, 0)$, when $k \neq 0$.

Question 3.2.5. Classify all invariant symmetric bilinear forms for $A_1{}^{(1)}$.

Question 3.2.6. Compute the cardinality $\|P_+^k\|$ for all series $A_r{}^{(1)}, B_r{}^{(1)}, C_r{}^{(1)}, D_r{}^{(1)}$. (*Hint*: this can always be done using one or two binomial coefficients.)

Question 3.2.7. The affine Weyl group of $A_1{}^{(1)}$ has two generators, which we call here ω and t. These act on $\mathbb{Z}^2$ as follows:

$$\omega(a, b) = (-a, b + 2a), \qquad t(a, b) = (3a + 2b, -2a - b).$$

(a) Find a formula for the action of t^n on (a, b). Find the orders of ω and t, and the determinants $\det(\omega)$ and $\det(t)$.
(b) Let $\beta = (a, b) \in \mathbb{Z}^2$ obey $k := a + b > 0$. Write $\rho = (1, 1)$. Show that the affine Weyl orbit of $\beta + \rho$ intersects

$$P_{++}^{k+2} := \{(1, k + 1), (2, k), \ldots, (k, 2), (k + 1, 1)\}$$

in at most one point, and that the orbit fails to intersect P_{++}^{k+2} iff $\beta + \rho$ is fixed by some nontrivial element of the affine Weyl group.

3.3 Generalisations of the affine algebras

Affine algebras are fascinating because they draw together so many different areas of mathematics and physics. Like anything else, they embed into assorted families in plenty of ways, each embedding preserving some properties and losing others. But do they embed into a much larger family of algebras that are also of interest outside Lie theory?

Generalisation is not the point of mathematics, and in fact, one must be honest, is usually rather dry. The challenge is to generalise in a rich and revealing direction. One of the more reliable ways of doing this is *closure*. Suppose we like to perform a certain activity, which unfortunately sometimes results in our toys being flung from our sandbox.

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

Then we build a bigger sandbox. When we divide integers, we don't always get integers, so we construct the rationals. When we take limits of rationals, we don't always get rationals, so we construct the reals. When we take square-roots of reals, we don't always get reals, so we construct the complex numbers.

Another appealing strategy for generalisation – *analogy* – was followed by Moody at the birth of Kac–Moody algebras (Section 3.2.1). However this strategy, even in the hands of a master, will not always be successful. This section reviews various generalisations of affine algebras, all obtained through analogy. Most important for our story are the Borcherds–Kac–Moody algebras, which have played a key role for instance in the proof of the Monstrous Moonshine conjectures.

3.3.1 Kac–Moody algebras

Recall the presentation R1, R2 of simple Lie algebras given in Definition 1.4.5, defined in terms of a Cartan matrix C1–C4. From the point of view of generators and relations, the step from 'finite-dimensional simple' to 'Kac–Moody' is rather easy: the only difference is that we drop the 'positive-definite' condition C4 (which was responsible for finite-dimensionality). That is:

Definition 3.3.1 (a) *A* Cartan$_{KM}$ matrix *A is any* $\ell \times \ell$ *integral matrix A obeying* C1, C2, C3 *(see Definition 1.4.5(a)), together with*

C4$'$ *there exists a positive diagonal matrix D such that the product AD is symmetric (i.e.* $(AD)^t = AD$*).*

(b) *Given any Cartan$_{KM}$ matrix A, the* Kac–Moody algebra $\mathfrak{g} = \mathfrak{g}(A)$ *is the Lie algebra with generators* e_i, f_i, h_i*, subject as before to the relations* R1 *and* R2 *(see Definition 1.4.5(b)).*

What we call Kac–Moody algebras are usually called *symmetrisable* Kac–Moody algebras in the literature. The adjective 'symmetrisable' emphasises the requirement C4$'$, which we shall always assume; dropping it means losing the invariant bilinear form, among other things. What we call 'Cartan$_{KM}$ matrix' here is usually called 'generalised symmetrisable Cartan matrix', but although that use of the word 'generalised' is traditional, it is now inappropriate (see Definition 3.3.4 below). More generally, appending 'generalised' to a term is an unimaginative empty cop-out that should be banned.

The theory of Kac–Moody algebras is quite parallel to that of the finite-dimensional simple Lie algebras. They are also generated by (finitely many) A_1 subalgebras. Most entries of A again are zero, so it is most convenient to graphically represent A using the Coxeter–Dynkin diagram (recall their definition in Section 1.4.3). As before, we may without loss of generality take the Cartan$_{KM}$ matrices to be indecomposable (i.e. consider connected diagrams).

Lemma 3.3.2 ([328], section 4.3) *Let A be an indecomposable Cartan$_{KM}$ matrix. Then exactly one of the following possibilities holds:*
(Fin) $\det(A) \neq 0$ *– there exists a column vector* $u > 0$ *such that* $Au > 0$;

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

(Aff) the nullspace (i.e. 0-eigenspace) of A is one-dimensional – there is a column vector $u > 0$ such that $Au = 0$;
(Hyp) there is a column vector $u > 0$ such that $Au < 0$.

If the Cartan$_{KM}$ matrix A is of finite type, then the corresponding Lie algebra $\mathfrak{g}(A)$ is finite-dimensional and simple. If the matrix A is of affine type, then the algebra $\mathfrak{g}(A)$ is infinite-dimensional, but has a $\mathbb{Z}$-grading $\mathfrak{g}(A) = \sum_j \mathfrak{g}_j$ into finite-dimensional subspaces $\mathfrak{g}_j$ where dimensions $\dim(\mathfrak{g}_j)$ grow at most polynomially with j (see Section 3.2.1). The affine algebras come in two flavours – nontwisted and twisted – and are listed in Figures 3.2 and 3.3. For A of hyperbolic type, again $\mathfrak{g}(A)$ has a $\mathbb{Z}$-grading into finite-dimensional subspaces $\mathfrak{g}_j$, but their dimensions $\dim(\mathfrak{g}_j)$ grow exponentially with j. We are mostly interested in the nontwisted affine algebras (Section 3.2). Relatively little is known about the hyperbolic ones (but see Section 3.4.3).

The relation between the realisation in Section 3.2.2 of an affine algebra as a loop algebra and the presentation of Definition 3.3.1(b) is as follows. Consider for simplicity ${A_1}^{(1)}$. The relevant Cartan$_{KM}$ matrix is $A = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$; then $\mathfrak{g}(A) \cong \mathcal{L}_{poly}(A_1) \oplus \mathbb{C}C$, with the isomorphism identifying

$$e_1 \mapsto \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad f_1 \mapsto \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad h_1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$
$$e_0 \mapsto \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix}, \quad f_0 \mapsto \begin{pmatrix} 0 & 0 \\ t^{-1} & 0 \end{pmatrix}, \quad h_0 \mapsto C - \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

More generally, the central term C of the affine algebra is given by $C = \sum_i a_i^\vee h_i$. Note though that we are missing the derivation ℓ_0; we will return to that shortly.

For indecomposable A, $\mathfrak{g}(A)$ is simple iff the determinant $\det(A) \neq 0$. When $\det(A) = 0$, $\mathfrak{g}(A)$ has a centre of dimension $\ell - m$ where m is the rank of the matrix A.

The basic structure theorem for Kac–Moody algebras is:

Theorem 3.3.3 *Let $\mathfrak{g} = \mathfrak{g}(A)$ be a symmetrisable Kac–Moody algebra (over $\mathbb{R}$). Then:*

(a) *$\mathfrak{g}$ has triangular decomposition $\mathfrak{g} = \mathfrak{g}_+ \oplus \mathfrak{h} \oplus \mathfrak{g}_-$ where $\mathfrak{g}_+$ is the subalgebra generated by the e_i, $\mathfrak{g}_-$ is generated by the f_i and $\mathfrak{h} = \mathrm{span}\{h_i\}$ is the Cartan subalgebra;*
(b) *$\mathfrak{g}$ has a root space decomposition – formally calling e_i degree α_i and f_i degree $-\alpha_i$ and defining $\mathfrak{g}_\alpha$ to be the subspace of degree $\alpha \in \mathbb{Z}\alpha_1 + \mathbb{Z}\alpha_2 + \cdots$, we get $\mathfrak{h} = \mathfrak{g}_0$ and $\mathfrak{g}_\pm = \oplus_{\alpha \in \Delta_\pm} \mathfrak{g}_\alpha$, where $[\mathfrak{g}_\alpha, \mathfrak{g}_\beta] \subset \mathfrak{g}_{\alpha+\beta}$ and $\Delta_- = -\Delta_+$;*
(c) *there is an involution ω on $\mathfrak{g}$ for which $\omega e_i = f_i$, $\omega h_i = -h_i$ and $\omega \mathfrak{g}_\alpha = \mathfrak{g}_{-\alpha}$;*
(d) *$\dim \mathfrak{g}_\alpha < \infty$ and $\dim \mathfrak{g}_{\pm\alpha_i} = 1$;*
(e) *there is an invariant symmetric bilinear form $(\cdot|\cdot)$, that is $([ab]|c) = -(b|[ac])$, such that for each root $\alpha \neq 0$ the restriction of $(\cdot|\cdot)$ to $\mathfrak{g}_\alpha \times \mathfrak{g}_{-\alpha}$ is nondegenerate and $(\mathfrak{g}_\alpha|\mathfrak{g}_\beta) = 0$ whenever $\beta \neq -\alpha$;*
(f) *there is a linear assignment $\alpha \mapsto h_\alpha \in \mathfrak{h}$ such that for all $a \in \mathfrak{g}_\alpha$, $b \in \mathfrak{g}_{-\alpha}$ we have $[a, b] = (a|b)\, h_\alpha$.*

These α are called *roots* and the α_i *simple roots*, as before. The roots α can be regarded as linear functionals on $\mathfrak{h}$, in such a way that for any $x \in \mathfrak{g}_\alpha$ and $h \in \mathfrak{h}$, we have $[hx] =$

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

$\alpha(h)\,x$. The involution in (c) is the Cartan involution, and is needed in defining unitary representations. The bilinear form in (e) is the generalisation here of the Killing form. For simple roots α_i, h_{α_i} in (f) is h_i, and is sometimes denoted $\alpha_i^\vee$ and called a *co-root*. The field was taken to be $\mathbb{R}$ here for convenience (Question 3.3.1).

When $\det(A) = 0$, the bilinear form restricted to $\mathfrak{h}$ will be degenerate and the simple roots interpreted as linear functionals on $\mathfrak{h}$ will be linearly dependent. To get around this, extend the Cartan subalgebra by $\dim(\mathrm{Null}(A)) = \ell - m$ more vectors. Call $\mathfrak{h}^e$ the resulting $(2\ell - m)$-dimensional space. Extend the bilinear form to $\mathfrak{h}^e$ so that it becomes nondegenerate, and the domain of the simple roots $\alpha_i \in \mathfrak{h}^*$ to all of $\mathfrak{h}^e$ so they become linearly independent. Up to equivalence, there is a unique way to do this. The space $\mathfrak{g}(A)^e := \mathfrak{g}(A) + \mathfrak{h}^e$ is given a Lie algebra structure by extending the relations of Definition 3.3.1(b) to include

$$[hh'] = 0, \qquad \forall h, h' \in \mathfrak{h}^e, \tag{3.3.1a}$$

$$[he_i] = \alpha_i(h), \qquad \forall h \in \mathfrak{h}^e, \tag{3.3.1b}$$

$$[hf_i] = -\alpha_i(h), \qquad \forall h \in \mathfrak{h}^e. \tag{3.3.1c}$$

For a Cartan$_{KM}$ matrix A of affine type, $\mathfrak{g}(A)^e$ is isomorphic to the corresponding algebra $\mathfrak{g} = \overline{\mathfrak{g}}^{(N)}$ we defined in Section 3.2.2: the extra vector is the derivation ℓ_0. Whenever $\det(A) = 0$, $\mathfrak{g}(A)^e$ and not $\mathfrak{g}(A)$ is the correct algebra to consider. Write $\mathfrak{g}(A)^e := \mathfrak{g}(A)$ when $\det(A) \neq 0$. Theorem 3.3.3 holds for $\mathfrak{g}(A)^e$, provided $\mathfrak{h}$ there is replaced with $\mathfrak{h}^e$.

Unlike the finite-dimensional case, some root multiplicities $\mathrm{mult}(\alpha) := \dim \mathfrak{g}_\alpha$ may be > 1. The roots of $\mathfrak{g}(A)^e$ come in two flavours: *real* (with $(\alpha|\alpha) > 0$) and *imaginary* (with $(\alpha|\alpha) \leq 0$). The simple roots are all real. Real roots behave exactly like the roots of finite-dimensional $\mathfrak{g}$: for example, $\mathrm{mult}(\alpha) = 1$ and the only multiples of α that are also roots are $\pm\alpha$. Imaginary roots behave more like the nonroot $0 \in \mathfrak{h}^*$: for example, $\mathrm{mult}(\alpha) \geq 1$ and any multiple $\mathbb{Z}\alpha$ is also a root.

The Weyl group W here is generated by the reflections through the simple roots α_i, or equivalently by reflections through all real roots. It has the usual properties: for example, root multiplicities are constant within the W-orbits.

A Kac–Moody algebra $\mathfrak{g}(A)^e$ has all the familiar representation-theoretic definitions and properties. For any weights $\lambda \in \mathfrak{h}^{e*}$, Verma modules $M(\lambda)$ and the irreducible highest-weight module $L(\lambda)$ are defined as usual. In particular, highest-weight modules are spanned by vectors of the form

$$f_{i_m} f_{i_{m-1}} \cdots f_{i_1} v, \tag{3.3.2}$$

where v is the highest-weight vector. Weight-space decompositions hold as before, and characters $\mathrm{ch}_M(h)$ are defined as in (1.5.9a). The character of the Verma module $M(\lambda)$ again equals (3.2.6). Integrability is defined by the locally nilpotent condition (Section 3.2.3); again, $L(\lambda)$ is integrable iff all Dynkin labels $\lambda(h_i) \in \mathbb{N}$, iff $L(\lambda)$ is unitarisable. The character of an integrable $L(\lambda)$ is given by the *Weyl–Kac character formula*

$$\mathrm{ch}_{L(\lambda)} = \frac{\sum_{w \in W} \det(w)\, e^{w(\lambda+\rho)}}{e^\rho \prod_{\alpha>0} (1 - e^{-\alpha})^{\mathrm{mult}(\alpha)}}. \tag{3.3.3}$$

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

This is identical to the Weyl character formula (1.5.11), except that the sum and product are infinite, and the multiplicities of (imaginary) root spaces can be > 1. For affine algebras, it reduces to (3.2.11c).

Apart from the affine and finite-dimensional simple algebras, the other Kac–Moody algebras have yet to make a real impact on other areas of mathematics and mathematical physics. However, [**127**] and [**171**] anticipate that the hyperbolic Kac–Moody algebras E_{10} and E_{11} will appear in *M-theory*, the still-hypothetical physics underlying strings.

3.3.2 Borcherds' algebras

In his efforts to prove the Monstrous Moonshine conjectures, Borcherds further generalised affine algebras. It is easy to associate a Lie algebra to a matrix A, but which class of matrices will yield a deep theory? Borcherds found such a class by holding in his hand a single algebra – the fake Monster Lie algebra (Section 7.2.2) – which acted much like a Kac–Moody algebra, even though it had imaginary simple roots.

Definition 3.3.4 (a) *A* Cartan$_{BKM}$ matrix *A is a (possibly infinite) matrix $A = (a_{ij})$, $a_{ij} \in \mathbb{R}$, obeying*

GC1. *either $a_{ii} = 2$ or $a_{ii} \leq 0$;*
GC2. *$a_{ij} \leq 0$ for $i \neq j$, and $a_{ij} \in \mathbb{Z}$ when $a_{ii} = 2$; and*
GC3. *there is a diagonal matrix D with each $d_{ii} > 0$ such that DA is symmetric.*

(b) *The* universal Borcherds–Kac–Moody algebra $\widehat{\mathfrak{g}} = \widehat{\mathfrak{g}}(A)$ *is the Lie algebra with generators e_i, f_i, h_{ij}, subject to the relations* [**71**]:

GR1. $[e_i f_j] = h_{ij}$, $[h_{ij} e_k] = \delta_{ij} a_{ik} e_k$ *and* $[h_{ij} f_k] = -\delta_{ij} a_{ik} f_k$, *for all i, j;*
GR2. $(\mathrm{ad}\, e_i)^{1-a_{ij}} e_j = (\mathrm{ad}\, f_i)^{1-a_{ij}} f_j = 0$, *whenever both $a_{ii} = 2$ and $i \neq j$; and*
GR3. $[e_i e_j] = [f_i f_j] = 0$ *whenever* $a_{ij} = 0$.

As before, the adjective 'symmetrisable' is usually appended in the literature. Unfortunately, the name 'Borcherds' is often replaced with the abomination 'generalised'. Note that for each i, span$\{e_i, f_i, h_{ii}\}$ is isomorphic to $\mathfrak{sl}_2(\mathbb{C})$ when $a_{ii} \neq 0$ and to $\mathfrak{Heis}$ (recall (1.4.3)) when $a_{ii} = 0$. Immediate consequences of the definition are that: (i) $[h_{ij} h_{mn}] = 0$; (ii) $h_{ij} = 0$ unless the ith and jth column of A are identical; (iii) the h_{ij} for $i \neq j$ lie in the centre of $\widehat{\mathfrak{g}}$. Setting all $h_{ij} = 0$ for $i \neq j$ gives the definition of the *Borcherds–Kac–Moody algebra* $\mathfrak{g} = \mathfrak{g}(A)$ [**69**]. This central extension $\widehat{\mathfrak{g}}$ of $\mathfrak{g}$ is introduced for its role in Theorem 3.3.6 below. If A has no zero columns, then $\widehat{\mathfrak{g}}$ equals its own universal central extension [**71**]. Because a Borcherds–Kac–Moody algebra can satisfy fewer relations, it typically contains a large free Lie subalgebra [**323**] (a free Lie algebra is analogous to a free group).

A universal Borcherds–Kac–Moody algebra differs from a Kac–Moody algebra in that it is built up from Heisenberg algebras as well as A_1, and these subalgebras intertwine in more complicated ways. Nevertheless, *much of the theory for finite-dimensional simple Lie algebras continues to find an analogue in this much more general setting* (e.g.

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

root-space decomposition, Weyl group, character formula, ...). This unexpected feature is the point of Borcherds–Kac–Moody algebras.

To get a feel for these algebras, let us prove a few simple results concerning the h_{ij}. Note first that, using the above relations together with anti-associativity, we obtain $[h_{ij}h_{k\ell}] = \delta_{ij}(a_{jk} - a_{j\ell})h_{k\ell}$. Comparing this with $[h_{k\ell}h_{ij}] = -[h_{ij}h_{k\ell}]$, we see that bracket must always equal 0. Hence all h's pairwise commute, and $h_{ij} = 0$ unless the ith and jth columns of A are identical.

The basic structure theorem is that of Kac–Moody algebras (Theorem 3.3.3):

Theorem 3.3.5 [**69**] *Let $\mathfrak{g} = \mathfrak{g}(A)$ be a Borcherds–Kac–Moody algebra (over $\mathbb{R}$). Then:*

(a) *$\mathfrak{g}$ has triangular decomposition $\mathfrak{g} = \mathfrak{g}_+ \oplus \mathfrak{h} \oplus \mathfrak{g}_-$ where $\mathfrak{g}_+$ is the subalgebra generated by the e_i, $\mathfrak{g}_-$ is generated by the f_i and $\mathfrak{h} = \mathrm{span}\{h_i\}$ is the Cartan subalgebra*;

(b) *$\mathfrak{g}$ has a root space decomposition – formally calling e_i degree α_i and f_i degree $-\alpha_i$, and defining $\mathfrak{g}_\alpha$ to be the subspace of degree $\alpha \in \mathbb{Z}\alpha_1 + \mathbb{Z}\alpha_2 + \cdots$, we get $\mathfrak{h} = \mathfrak{g}_0$ and $\mathfrak{g}_\pm = \oplus_{\alpha\in\Delta_\pm}\mathfrak{g}_\alpha$, where $[\mathfrak{g}_\alpha, \mathfrak{g}_\beta] \subset \mathfrak{g}_{\alpha+\beta}$ and $\Delta_- = -\Delta_+$*;

(c) *there is an involution ω on $\mathfrak{g}$ for which $\omega e_i = f_i$, $\omega h_i = -h_i$ and $\omega\mathfrak{g}_\alpha = \mathfrak{g}_{-\alpha}$*;

(d) *$\dim \mathfrak{g}_\alpha < \infty$ and $\dim \mathfrak{g}_{\pm\alpha_i} = 1$*;

(e) *there is an invariant symmetric bilinear form $(\cdot|\cdot)$ such that for each root $\alpha \neq 0$, the restriction of $(\cdot|\cdot)$ to $\mathfrak{g}_\alpha \times \mathfrak{g}_{-\alpha}$ is nondegenerate and $(\mathfrak{g}_\alpha|\mathfrak{g}_\beta) = 0$ whenever $\beta \neq -\alpha$*;

(f) *there is a linear assignment $\alpha \mapsto h_\alpha \in \mathfrak{h}$ such that for all $a \in \mathfrak{g}_\alpha$, $b \in \mathfrak{g}_{-\alpha}$, we have $[a, b] = (a|b)\, h_\alpha$.*

The condition that $\mathfrak{g}$ be symmetrisable (i.e. condition GC3) is necessary for the existence of the bilinear form in Theorem 3.3.5(e). As in Section 3.3.1, it is common to add derivations. In particular, define $D_i(a) = n_i a$ for any $a \in \mathfrak{g}_{n_1\alpha_1+\cdots}$; then each linear map D_i is a derivation, and adjoining these to $\mathfrak{h}$ defines an abelian algebra $\mathfrak{h}^e$. The simple root α_i can be interpreted as the element of $\mathfrak{h}^{e*}$ obeying $\alpha_j(h_i) = a_{ij}$ and $\alpha_j(D_i) = \delta_{ij}$. The role of the derivations is to make these simple roots linearly independent. Construct the induced bilinear form $(\cdot|\cdot)$ on $\mathfrak{h}^{e*}$, obeying $(\alpha_i \,|\, \alpha_j) = d_i a_{ij}$ (see [**322**] for details).

The properties in Theorem 3.3.5 characterise Borcherds–Kac–Moody algebras (see e.g. [**72**] for a proof):

Theorem 3.3.6 *Let L be a Lie algebra (over $\mathbb{R}$) satisfying the following conditions:*

(i) *L has a $\mathbb{Z}$-grading $\oplus_i L_i$, and $\dim L_i < \infty$ for all $i \neq 0$;*

(ii) *L has an involution ω sending L_i to L_{-i} and acting as -1 on L_0;*

(iii) *L has a contravariant bilinear form $(\cdot|\cdot)$ such that $(L_i|L_j) = 0$ if $i \neq -j$, and such that $-(a \,|\, \omega(a)) > 0$ if $0 \neq a \in L_i$ for $i \neq 0$.*

Then there is a homomorphism π from some $\widehat{\mathfrak{g}}(A)$ to L whose kernel is contained in the centre of $\widehat{\mathfrak{g}}(A)$, and L is the semi-direct product of the image of π with a subalgebra of the abelian subalgebra L_0. That is, L is obtained from $\widehat{\mathfrak{g}}$ by modding out some of the centre and adding some commuting derivations.

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

Conversely, any (real) Borcherds–Kac–Moody algebra obeys conditions (i), (ii) and (iii). For example, let $L = \mathfrak{sl}_2(\mathbb{R})$ and recall (1.4.2b). Then L has $\mathbb{Z}$-grading $L_{-1} \oplus L_0 \oplus L_1 = \mathbb{C}e \oplus \mathbb{C}h \oplus \mathbb{C}f$, $\omega(x) = -x^t$ and $(x|y) = \mathrm{tr}(xy)$. Theorem 3.3.6 tells us that *Borcherds–Kac–Moody algebras are the ultimate generalisation of simple Lie algebras,* in the sense that any further generalisation will lose some basic structural ingredient.

Let Π^{re} be the set of all *real simple roots*, i.e. all α_i with $a_{ii} = 2$; the remainder are the *imaginary simple roots* $\alpha \in \Pi^{im}$. The *Weyl group* W of $\mathfrak{g}^e$ is generated by the reflections $r_{\alpha_i} : \mathfrak{h}^{e*} \to \mathfrak{h}^{e*}$ for each $\alpha_i \in \Pi^{re}$: $r_i(\lambda) = \lambda - \lambda(h_i)\,\alpha_i$. It is a crystallographic Coxeter group (Section 3.2.1). The *real roots* of $\mathfrak{g}^e$ are defined to be those in $W(\Pi^{re})$; all other roots are called *imaginary*. For all real roots, $\dim\,(\mathfrak{g}^e)^\alpha = 1$ and $(\alpha\,|\,\alpha) > 0$.

Integrable highest-weight modules are defined as before: namely, each e_α, f_α must act locally nilpotently for all real roots α. More precisely, $V = \oplus_{\mu\in\mathfrak{h}^{e*}} V_\mu$ where the weight-space $V_\mu := \{v \in V \mid h.v = \mu(h)v\}$, with $\dim V_\mu < \infty$, and whenever $a_{ii} = 2$, $(e_i)^k.v = 0 = (f_i)^k.v$ for all $v \in V$ and all sufficiently large k. By the *character* we mean the formal sum $\mathrm{ch}_V := \sum_{\mu\in\mathfrak{h}^{e*}} (\dim V_\mu)\, e^\mu$. Let P_+ be the set of all weights $\lambda \in \mathfrak{h}^{e*}$ obeying $\lambda(h_i) \in \mathbb{N}$ whenever $a_{ii} = 2$, and $\lambda(h_i) \geq 0$ for all other i. Define the highest-weight $\mathfrak{g}^e$-module $L(\lambda)$ in the usual way as the quotient of the Verma module by the largest proper graded submodule. Choose $\rho \in \mathfrak{h}^{e*}$ to satisfy $(\rho\,|\,\alpha_i) = \frac{1}{2}(\alpha_i\,|\,\alpha_i)$ for all i, and define $S_\lambda = e^{\lambda+\rho}\sum_s \epsilon(s)\, e^s$ where s runs over all sums of imaginary simple roots and $\epsilon(s) = (-1)^m$ if s is the sum of m distinct mutually orthogonal imaginary simple roots, each of which is orthogonal to λ, otherwise $\epsilon(s) = 0$. Then we get the *Weyl–Kac–Borcherds character formula*:

$$\mathrm{ch}_{L(\lambda)} = \frac{\sum_{w\in W} \epsilon(w)\, w(S_\lambda)}{e^\rho \prod_{\alpha\in\Delta_+}(1 - e^{-\alpha})^{\mathrm{mult}\,\alpha}} \tag{3.3.4}$$

(compare (3.3.3)). S_λ is the correction factor due to imaginary simple roots.

Thus Borcherds' algebras strongly resemble Kac–Moody ones and constitute a natural and nontrivial generalisation. The main differences are that they can be generated by copies of the Heisenberg algebra as well as $\mathfrak{sl}_2(\mathbb{R})$, and that there can be imaginary simple roots. For more on their theory, see, for example, [**328**] chapter 11.13, [**272**], [**322**], [**469**]. Interesting examples are the Monster Lie algebra (Section 7.2.2), whose (twisted) denominator identity supplied the relations needed to complete the proof of the Monstrous Moonshine conjectures, and the fake Monster [**70**]. A Borcherds–Kac–Moody algebra can be associated with any even Lorentzian lattice, and also with any Calabi–Yau manifold [**275**]. Of course it is a broad enough class that almost all of them will be uninteresting; an intriguing approach to identifying the interesting ones is sketched at the end of Section 3.4.3.

We know simple Lie algebras arise in both classical and quantum physics, and the affine Kac–Moody algebras are important in conformal field theory, as we see next chapter. Borcherds–Kac–Moody algebras have appeared in the physics literature in the context of BPS states in string theory (see [**275**]), and as a possible symmetry of M-theory [**285**].

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

3.3.3 Toroidal algebras

As mentioned in Section 3.2.6, replacing the loop algebra $S^1 \to \overline{\mathfrak{g}}$ with more general spaces $M \to \overline{\mathfrak{g}}$ has a very different theory and seems much more complicated. The most obvious generalisation of affine algebras, which has a chance of retaining some of their special properties, is to replace the loop algebra $S^1 \to \overline{\mathfrak{g}}$ with a space of maps $S^1 \times \cdots \times S^1 \to \overline{\mathfrak{g}}$. As $S^1 \times \cdots \times S^1$ (n times) is topologically the n-dimensional torus, these are called *toroidal algebras*. We will try to mimic the theory of loop algebras as far as we can. If nothing else, we will identify some features responsible for making the earlier theory so special.

Let $\overline{\mathfrak{g}}$ be a simple finite-dimensional Lie algebra. Choose any $n \geq 1$, and let $\widetilde{\mathfrak{g}}$ be the *multi-loop algebra*, i.e. tensor product $\overline{\mathfrak{g}} \otimes \mathbb{C}[t_0^{\pm 1}, \ldots, t_n^{\pm 1}]$ of $\overline{\mathfrak{g}}$ with Laurent polynomials in formal variables t_i. Then $\widetilde{\mathfrak{g}}$ is a Lie algebra with $\mathbb{Z}^{n+1}$-grading into finite-dimensional subspaces. The following theory treats as distinguished one of these $n+1$ variables, namely t_0. To complete the construction of the toroidal algebra, we take the universal central extension $0 \to \mathcal{K} \to \widetilde{\mathfrak{g}} \otimes \mathcal{K} \to \widetilde{\mathfrak{g}} \to 0$ of the multi-loop algebra $\widetilde{\mathfrak{g}}$, and then adjoin sufficiently many derivations (as we've done throughout this chapter). However, both of these extensions are infinite-dimensional. More precisely, write $d_i = t_i \, \mathrm{d}/\mathrm{d}t_i$ for the degree-derivation for variable t_i. Let $\mathcal{D}^*$ denote the algebra of derivations $\oplus_{i=1}^n \mathbb{C}[t_0^{\pm 1}, \ldots, t_n^{\pm 1}]d_i \oplus \mathbb{C}d_0$. The resulting Lie algebra structure on the space $\widetilde{\mathfrak{g}} \oplus \mathcal{K} \oplus \mathcal{D}^*$ is uniquely determined up to a 2-cocycle $\tau : \mathcal{D}^* \times \mathcal{D}^* \to \mathcal{K}$, which defines how the bracket of derivations contributes a central term. There is a two-dimensional space of these τ; choosing any of them defines a *toroidal Lie algebra* $\mathfrak{g}_\tau$. Adding $\mathcal{D}^*$ reduces the centre from the infinite-dimensional $\mathcal{K}$ to an $(n+1)$-dimensional space. See [**53**] for more details of the construction of $\mathfrak{g}_\tau$.

The role of the Virasoro algebra (which as we know is a central extension of $\mathrm{Der}(\mathbb{C}[t^{\pm 1}]) = \mathrm{Vect}(S^1) \otimes \mathbb{C})$ is here replaced by an abelian extension [**173**] of the complex vector fields on a torus or equivalently of $\mathrm{Der}(\mathbb{C}[t_0^{\pm 1}, \ldots, t_n^{\pm 1}])$. It is a Lie algebra $\mathfrak{V}_\tau$ parametrised by the 2-cocycle τ, defined on the space $\mathcal{K} \oplus \mathrm{Der}(\mathbb{C}[t_0^{\pm 1}, \ldots, t_n^{\pm 1}])$. $\mathfrak{V}_\tau$ acts for instance on the Verma modules of $\mathfrak{g}_\tau$. We will be more interested in the Lie subalgebra $\mathfrak{v}_\tau = \mathcal{K} \oplus \mathcal{D}^*$ of $\mathfrak{g}_\tau$. The modules constructed below carry a projective action of the Witt algebra $\mathbb{C}[t_0^{\pm 1}]d_0$, as in the affine setting.

Affine algebras exist for their (integrable) modules and in particular their characters, so we need to find an interesting class of modules for the toroidal algebras. This isn't easy to do, but major progress was made in [**53**]. Let L_τ be an irreducible highest-weight module of level $k \neq 0$, for the affine algebra $\overline{\mathfrak{g}}^{(1)}$, and let W be any finite-dimensional module for $\mathfrak{gl}_N$. Then [**53**] constructs an irreducible $\mathfrak{g}_\tau$-module $M_{\lambda,W}$ that has finite-dimensional homogeneous spaces with respect to the natural $\mathbb{Z}^{n+1}$-grading, and thus has a character. More precisely, they first obtain a $\mathfrak{v}_\tau$-module by applying a Verma-like construction to $W \otimes \mathbb{C}[t_1^{\pm 1}, \ldots, t_n^{\pm 1}]$, and then they take the irreducible quotient M_W as usual; finally, they define a $\mathfrak{g}_\tau$-module structure on the tensor product $M_{\lambda,W} := L_\lambda \otimes M_W$. In [**54**] they show that these are modules of a 'near-vertex operator algebra' (see Definition 5.1.3(c)) closely related to affine algebra vertex operator algebras at generic level. From this,

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

their characters can be computed, and familiar modular forms arise. This is promising because interesting Lie algebra modules seem to be the ones that arise as modules of related structures (e.g. Lie groups or vertex operator algebras). It is too easy to be a Lie algebra module. On the other hand, these are surely not the best $\mathfrak{g}_\tau$-modules – they have only found the analogue of generic $L(\lambda)$, but not yet the analogue of the 'integrable' modules. Their characters are like (3.1.7a), but we would like to identify modules with characters analogous to the discrete series. By analogy with better-understood algebras, we should look for modules with maximal numbers of 'null vectors' quotiented out.

It may seem artificial to choose a distinguished direction (namely the 0th), but to some extent this is inevitable. It is an elementary consequence of Schur's Lemma (recall Lemma 1.1.3) that in these irreducible $\mathfrak{g}_\tau$-modules, the centre span$\{K_0, \ldots, K_n\}$ should act as scalars, and thus an n-dimensional subspace must act trivially. These representations are designed so that K_0 is nontrivial but the other K_i act trivially.

What is natural to pursue from, for example, an algebraic point of view, and what is a successful theory from that point of view, is not necessarily of more general interest. It is from this broader, multidisciplinary standpoint that we (unfairly) judge the value of these generalisations. There is a large class of $\mathfrak{g}_\tau$-modules (namely those described above) whose characters have (fairly weak) modularity properties, but this seems to arise solely from the well-milled Heisenberg algebra combinatorics and it isn't clear yet that they have independent value. Possible physical relevance in Wess–Zumino–Witten models in more than two space-time dimensions is explored in, for example, [**306**]. The jury is still out on the greater relevance of toroidal algebras to, for example, Moonshine or physics, and certainly more work is needed.

3.3.4 Lie algebras and Riemann surfaces

The previous subsection emphasises the difficulties of higher-dimensional analogues of loop algebras. Perhaps the best generalisation of the affine algebras, particularly in the sense of retaining and enriching automorphic properties of the characters, associates infinite-dimensional Lie algebras to each Riemann surface with marked points. This theory has been developed in a series of papers by Krichever–Novikov, Bremner, Schlichenmaier, Sheinman and others – see [**491**] for a list of references. The starting point is a reinterpretation of the Laurent polynomials $\sum a_n t^n \in \mathcal{L}_{poly}\overline{\mathfrak{g}}$. Before, we interpreted the formal variable t as a point on the unit circle $S^1 \subset \mathbb{C}$, but now we regard t as lying in the punctured plane $\mathbb{C}\backslash\{0\}$, or equivalently the twice-punctured Riemann sphere $\mathbb{P}'(\mathbb{C})$. Similarly, the Witt algebra Vect(S^1) can be interpreted as the Lie algebra of meromorphic vector fields on $\mathbb{P}'(\mathbb{C})$ with possible poles only at 0 and ∞.

Let Σ be any Riemann surface of genus g, and choose $p > 1$ distinct ordered points $P = (z_1, \ldots, z_p)$, $z_i \in \Sigma$. In the language of string theory described next chapter, we can think of Σ as being a world-sheet corresponding to p asymptotic incoming or outgoing strings (Section 4.3.1). Let $\mathcal{A}_{\Sigma,P}$ be the space of functions meromorphic on Σ, with possible poles only at P, and let $\mathcal{L}_{\Sigma,P}$ be the space of meromorphic vector fields on Σ, again with possible poles only at P. The bracket of $\mathcal{L}_{\Sigma,P}$ comes from the Lie

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

derivative, as usual with vector fields, while the bracket of $\mathcal{A}_{\Sigma,P}$ is taken to be trivial. Let $\overline{\mathfrak{g}}$ be any simple finite-dimensional Lie algebra. The loop algebra $\mathcal{L}_{poly}\overline{\mathfrak{g}}$ is replaced with $\mathfrak{g}_{\Sigma,P} := \mathcal{A}_{\Sigma,P} \otimes \overline{\mathfrak{g}}$, with bracket $[\sum_i f_i \otimes x_i, \sum_j g_j \otimes y_j] = \sum_{i,j} f_i g_j [x_i y_j]$. The Laurent polynomials $\mathbb{C}[t^{\pm 1}]$ are replaced with $\mathcal{A}_{\Sigma,P}$. The Witt algebra is replaced with $\mathcal{L}_{\Sigma,P}$. Just as $\mathfrak{Witt}$ acts on $\mathcal{L}_{poly}\overline{\mathfrak{g}}$ by derivations, so does $\mathcal{L}_{\Sigma,P}$ act on $\mathfrak{g}_{\Sigma,P}$.

There are some subtle differences with the more familiar loop algebras. The loop algebras have an important $\mathbb{Z}$-grading. These higher-genus algebras $\mathcal{L}_{\Sigma,P}$ and $\mathfrak{g}_{\Sigma,P}$ have instead an *almost-grading* by $\mathbb{Z}$, in the sense that $\mathcal{L}_{\Sigma,P}$ (say) can be decomposed $\mathcal{L}_{\Sigma,P} = \oplus(\mathcal{L}_{\Sigma,P})_n$ as a vector space into finite-dimensional subspaces $(\mathcal{L}_{\Sigma,P})_n$, such that

$$[(\mathcal{L}_{\Sigma,P})_m, (\mathcal{L}_{\Sigma,P})_n] \subseteq \oplus_{\ell=m+n+L}^{m+n+M} (\mathcal{L}_{\Sigma,P})_\ell$$

for some fixed integers $L, M \in \mathbb{Z}$. This would be a true grading if $M = L = 0$. The algebra $\mathfrak{g}_{\Sigma,P}$ behaves similarly. The subspaces $(\mathcal{L}_{\Sigma,P})_n$ and $(\mathfrak{g}_{\Sigma,P})_n$ are defined by considering orders of poles (and splitting P into incoming and outgoing points).

In the loop algebra situation, for $\overline{\mathfrak{g}}$ simple, there is a unique nontrivial central extension. On the other hand, $\mathfrak{g}_{\Sigma,P}$ typically has several. However, only one will be compatible with the almost-grading, and so that is the one we choose. Call it $\widehat{\mathfrak{g}_{\Sigma,P}}$. Similarly, we get a unique central extension $\widehat{\mathcal{L}_{\Sigma,P}}$ of $\mathcal{L}_{\Sigma,P}$, which in the special case of a sphere with one incoming and one outgoing puncture is $\mathfrak{Vir}$.

Verma modules, etc. for $\widehat{\mathfrak{g}_{\Sigma,P}}$ can be defined as before using the universal enveloping algebra, and are parametrised by $p = \|P\|$ highest weights $\lambda^{(1)}, \ldots, \lambda^{(p)} \in \mathfrak{h}^*$ and a complex number k (the level). For these modules $W_{(\lambda,k)}$, $\lambda = (\lambda^{(1)}, \ldots, \lambda^{(p)})$, there is an analogue of the Sugawara construction (3.2.15), which shows that each of these $\widehat{\mathfrak{g}_{\Sigma,P}}$-modules $W_{(\lambda,k)}$ is simultaneously a $\widehat{\mathcal{L}_{\Sigma,P}}$-module, in perfect analogy with the affine situation.

Physically, these algebras $\widehat{\mathfrak{g}_{\Sigma,P}}$ and $\widehat{\mathcal{L}_{\Sigma,P}}$ should be regarded as higher-genus global symmetries for, for example, the Wess–Zumino–Witten models discussed next chapter. Locally, that is in terms of local coordinates at each marked point z_i, we get a copy of the affine algebra $\overline{\mathfrak{g}}^{(1)}$ and Virasoro algebra $\mathfrak{Vir}$. A module for, for example, $\widehat{\mathfrak{g}_{\Sigma,P}}$ similarly specialises to the $\overline{\mathfrak{g}}^{(1)}$-module $L(\lambda^{(i)})$ at each point $z_i \in P$.

The theory is still a work in progress – see, for example, [**491**], [**492**] and references therein. But it can be expected that for each positive level k and choice of Σ, and p highest weights $\lambda^{(i)} \in P_+^k(\mathfrak{g})$, a number of level-$k$ representations of $\widehat{\mathfrak{g}_{\Sigma,P}}$ will be singled out (the exact number being given by Verlinde's formula (6.1.2)), and these will 'transform covariantly' with respect to the mapping class group of $\Sigma \backslash P$. Obviously this is an exciting direction that should be pursued, with direct relevance to higher-genus Moonshine (Section 6.3.1).

Question 3.3.1. (a) Define $D = \prod_{\alpha \in \Delta_+} (1 - e^{-\alpha})^{\text{mult}(\alpha)}$. Verify $r_i(D) = e^{-\alpha_i} D$.
(a) Find a vector $r \in \mathfrak{h}$ such that $w(e^r D) = \epsilon(w)\, e^r D$.

Question 3.3.2. Let A be a Cartan$_{BKM}$ matrix, and $\mathfrak{g}$ the corresponding universal Borcherds–Kac–Moody algebra.
(a) Prove h_{ij} lies in the centre.

(b) Suppose the ith and jth rows of A are identical. Then show that $h_{ii} - h_{jj}$ is in the centre of $\mathfrak{g}$.

Question 3.3.3. In what ways (if any) do Theorems 3.3.3, 3.3.5, 3.3.6 change if the field is $\mathbb{C}$ and not $\mathbb{R}$?

Question 3.3.4. Prove that for any Lie algebra L obeying conditions (i), (ii), (iii) of Theorem 3.3.6, L_0 will be an abelian subalgebra.

3.4 Variations on a theme of character

3.4.1 Twisted #3: twisted representations

In this subsection we complete the introduction of the twisted character which we began in Section 1.5.4. These are to the usual character what the McKay–Thompson series are to the j-function. In Section 5.3.6 we generalise this construction, but as always the special case of affine algebras is particularly pretty and significant. The reader is encouraged to reread Section 1.5.4 for background.

Let's start with a twisted affine algebra $\overline{\mathfrak{g}}^{(N)}$, obtained as in (3.2.4) from the nontwisted algebra $\mathfrak{g} = \overline{\mathfrak{g}}^{(1)}$ and an order-N symmetry α of the Coxeter–Dynkin diagram of $\overline{\mathfrak{g}}$. Consider any integrable highest-weight $\overline{\mathfrak{g}}^{(N)}$-module $L(\lambda)$, $\lambda \in P_+^k(\overline{\mathfrak{g}}^{(N)})$. Think of this as a representation ρ. We can extend ρ linearly to $\mathfrak{g}$, by defining

$$\rho(xt^n) = \xi_N^{i-n}\rho(xt^n), \tag{3.4.1a}$$

for x in the α-eigenspace $(\overline{\mathfrak{g}})_i$ (Section 1.5.4). This isn't a true representation of $\mathfrak{g}$ – it's called a *twisted representation* of $\mathfrak{g}$, as it obeys

$$[\rho(xt^n), \rho(yt^m)] = \xi_N^{i+j-n-m}\rho([xt^n, yt^m]), \tag{3.4.1b}$$

when $x \in (\overline{\mathfrak{g}})_i$ and $y \in (\overline{\mathfrak{g}})_j$. Thus a true representation of the twisted affine algebra $\overline{\mathfrak{g}}^{(N)}$ corresponds to a twisted representation of the nontwisted algebra $\overline{\mathfrak{g}}^{(1)}$. In Section 5.4.6 we extend this notion of twisted representation to vertex operator algebras.

Twisted representations are vaguely reminiscent of projective representations. But a projective representation becomes a true representation when the algebra is extended, while a twisted representation becomes a true representation when the algebra is shrunk. Groups most naturally have projective representations, vertex operator algebras most naturally have twisted ones, and affine algebras have both.

Consider more generally any symmetry α of the Coxeter–Dynkin diagram of $\mathfrak{g}$. As in Section 3.2.2, α extends to an automorphism of $\mathfrak{g}$ (e.g. $\alpha(e_i) = e_{\alpha i}$, and α fixes the centre and derivation). Because of this, α permutes the $\mathfrak{g}$-modules as in Section 1.5.4. In particular, α takes the highest-weight module $L(\lambda)$ to $L(\lambda^\alpha)$, where $(\lambda^\alpha)_i = \lambda_{\alpha i}$, and moreover takes weight-space $L(\lambda)_\mu$ to weight-space $L(\lambda^\alpha)_{\mu^\alpha}$. All of this generalises to any Borcherds–Kac–Moody algebra.

Now suppose $\lambda^\alpha = \lambda$, that is λ is a fixed point of α. Then $L(\lambda)$ and $L(\lambda)^\alpha$ are isomorphic as $\mathfrak{g}$-modules, so let τ_α be a linear isomorphism of the space $L(\lambda)$ that intertwines their $\mathfrak{g}$-actions: that is, $\alpha(x).v = x.\tau_\alpha(v)$ in terms of the $\mathfrak{g}$-action of $L(\lambda)$. Because $L(\lambda)$ is irreducible, τ_α is uniquely determined up to a scalar multiple; scaled appropriately, it will permute all vectors of the form (3.3.2). By the α*-twisted character* or *twining character* χ_λ^α we mean

$$\chi_\lambda^\alpha(h) = \exp\left[\left(-h_\lambda + \frac{c_\lambda}{24} + \frac{k\,(d-d^{orb})}{24h^\vee}\right)\delta\right] \mathrm{tr}_{L(\lambda)}\tau_\alpha e^h$$
$$= \exp\left[\left(-h_\lambda + \frac{c_\lambda}{24} + \frac{k\,(d-d^{orb})}{24h^\vee}\right)\delta\right] \sum_{\mu=\alpha\mu} \mathrm{tr}(\tau_\alpha)e^{\mu(h)}, \quad \forall h\in\mathfrak{h} \quad (3.4.2a)$$

where d and d^{orb} are the dimensions of the semi-simple Lie algebras $\overline{\mathfrak{g}}$ and $\overline{\mathfrak{g}^{orb}}$ (the algebra $\mathfrak{g}^{orb}$ is defined in Theorem 3.4.1) and h_λ, c_λ are in (3.2.9). As in (3.2.9a), the normalisation here is chosen to make modularity simplest – see (3.4.2b) below. As we see from (3.2.5d), the vector $\delta\in\mathfrak{h}^*$ in (3.4.2a) isolates the coefficient $2\pi i\tau$ of the derivation ℓ_0.

Theorem 3.4.1 [**213**] *Let $\mathfrak{g} = X_r{}^{(1)}$ be a nontwisted affine algebra, and let α be a symmetry of the Coxeter–Dynkin diagram of $\mathfrak{g}$. Then for any integrable highest-weight λ of $\mathfrak{g}$, with $\alpha\lambda=\lambda$, the α-twisted character $\chi_\lambda^\alpha(h)$, restricted to any $h\in\mathfrak{h}$ fixed by α, equals some true character $\chi_{\widetilde{\lambda}}(h)$ of the 'orbit Lie algebra' $\mathfrak{g}^{orb} = ((\mathfrak{g}^{op})_0)^{op}$.*

'$\mathfrak{g}^{op}$' is the affine Kac–Moody algebra whose Coxeter–Dynkin diagram is that of $\mathfrak{g}$ except with all arrows reversed. Note that $\mathfrak{g}^{orb}$ is not a subalgebra of $\mathfrak{g}$, although its Cartan subalgebra $\mathfrak{h}^{orb}$ can be identified with that $\mathfrak{h}_0$ of the fixed-point subalgebra $\mathfrak{g}_0$. What is special about $\mathfrak{g}^{orb}$ is that there is a natural map P_α (see Section 3.3 of [**213**] for its precise construction) sending $\mathfrak{g}$-weights fixed by α to the weights of $\mathfrak{g}^{orb}$, and preserving all inner-products. The weight $\widetilde{\lambda}$ in Theorem 3.4.1 is $P_\alpha(\lambda)$. The normalisation in (3.4.2a) is exactly what one would expect for a character of $\mathfrak{g}^{orb}$:

$$h^{orb}_{\widetilde{\lambda}} - \frac{c^{orb}}{24} = h_\lambda - \frac{c}{24} + \frac{k\,(d-d^{orb})}{24h^\vee}. \quad (3.4.2b)$$

For example, consider $\mathfrak{g} = A_{2n-1}{}^{(1)}$ and $\mathfrak{g} = A_{2n}{}^{(1)}$, respectively, with α being the left–right reflection symmetry ('charge-conjugation') 'C' fixing the 0th node. Then the orbit Lie algebra $\mathfrak{g}^{orb}$ is the twisted affine algebras $D_{n+1}{}^{(2)}$ and $A_{2n}{}^{(2)}$, respectively. For $\mathfrak{g} = \mathfrak{sl}_n{}^{(1)}$ with a cyclic symmetry ('simple-current') '$J^{n/d}$' of order d (so d divides n), $\mathfrak{g}^{orb} = \mathfrak{sl}_{n/d}{}^{(1)}$. The map P_α in these examples is

$$P_C : \lambda_0\omega_0 + \sum_{i=1}^{n-1}\lambda_i\,(\omega_i+\omega_{2n-i}) + \lambda_n\omega_n \mapsto \sum_{j=0}^{n}\lambda_i\omega_i^{orb},$$
$$P_C : \lambda_0\omega_0 + \sum_{i=1}^{n}\lambda_i\,(\omega_i+\omega_{2n-i}) \mapsto \sum_{j=0}^{n}\lambda_i\omega_i^{orb},$$
$$P_{J^{n/d}} : \sum_{i=0}^{n/d-1}\lambda_i\left(\omega_i+\omega_{i+n/d}+\cdots+\omega_{i+n-n/d}\right) \mapsto \sum_{j=0}^{n/d-1}\lambda_i\omega_i^{orb}.$$

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

The map P_α is not mysterious. For example, for $\mathfrak{g} = \mathfrak{sl}_{2n}{}^{(1)}$ and $\alpha = C$, the fundamental weights ω_i^{orb} of $\mathfrak{g}^{orb}$ are the obvious basis for the C-invariant weights of $\mathfrak{g}$, namely $\omega_i^{orb} = \omega_i + \omega_{2n-i}$ (for $1 \leq i < n$) together with $\omega_0^{orb} = \omega_0$ and $\omega_n^{orb} = \omega_n$.

The most important case in Theorem 3.4.1 is the degenerate one. The Coxeter–Dynkin diagram of $\mathfrak{sl}_n{}^{(1)}$ has an order-n cyclic symmetry J. In this case, an α-fixed point looks like $\lambda = (\lambda_0, \lambda_0, \ldots, \lambda_0)$ for $\lambda_0 = k/n$, and the α-twisted character $\chi_\lambda^\alpha(h)$, restricted to h fixed by α, equals the τ-independent function $\exp[2\pi \mathrm{i}\,(\overline{\lambda}(\overline{h}) + ku)]$ – that is, only the top weight-space survives.

A good question in Lie theory is always rewarded with a beautiful answer. Theorem 3.4.1 holds more generally for any Borcherds–Kac–Moody algebra. The proof follows that of the Weyl–Kac–Borcherds character formula.

We get from Theorems 3.4.1 and 3.2.4 that the twisted characters are modular functions, and obey an analogue of Theorem 3.2.3. As an isolated example, this is rather surprising, but it fits into a much larger context (Section 5.3.6). We also find there how modular transformations relate the twisted characters to twisted representations – it is quite analogous to (2.3.10b). From this greater context of vertex operator algebra modules and characters twisted by automorphisms, the modularity of these twisted characters is not so surprising. What is more surprising is positivity, that is, the q-expansion has positive integer coefficients. This is true, for instance, for only two-thirds of the McKay–Thompson series T_g. See Section 7.3.5, especially Conjecture 7.3.3, for an analogous result for the Moonshine module $V^\natural$.

3.4.2 Denominator identities

A very useful formula for the characters of simple finite-dimensional Lie algebras $\overline{\mathfrak{g}}$ is the Weyl character formula (1.5.11). It is rare indeed when the trivial special case of a theorem or formula is interesting. But that happens here. Consider the trivial representation: i.e. $x \mapsto 0$ for all $x \in \overline{\mathfrak{g}}$. Then the character (1.5.9a) is identically 1: $\mathrm{ch}_0 \equiv 1$. Thus the character formula tells us that a certain alternating sum over the Weyl group W equals a certain product over positive roots $\alpha \in \Delta_+$:

$$\prod_{\alpha\in\Delta_+} \left(1 - e^{-\alpha(z)}\right) = e^{-\rho(z)} \sum_{w\in W} \epsilon(w)\, e^{w(\rho)(z)}. \tag{3.4.3}$$

Here, z lies in the Cartan subalgebra $\mathfrak{h}$, and the Weyl vector ρ is $\omega_1 + \cdots + \omega_r$. Equation (3.4.3) is called a *denominator identity*. For the smallest simple algebra A_1, (3.4.3) is trivial: $1 - e^{-z} = e^{-z/2}(e^{z/2} - e^{-z/2})$. For A_2 we get a sum of six terms equalling a product of three terms, and the complexity continues to rise from there.

In particular, look at $\overline{\mathfrak{g}} = \mathfrak{sl}_n(\mathbb{C})$. We can realise the roots, etc. of $\overline{\mathfrak{g}}$ in terms of an orthonormal basis $\{e_i\}$ of $\mathbb{C}^n$ as follows: the positive roots are $e_i - e_j$ for $1 \leq i < j \leq n$; the Cartan subalgebra $\mathfrak{h}$ is the hyperplane orthogonal to $\sum_i e_i$; the Weyl group is the symmetric group $\mathcal{S}_n$, acting on $\mathbb{C}^n$ and hence $\mathfrak{h}$ by permuting the e_i; the Weyl vector $\rho = \frac{1}{2}\sum_i (n+1-2i)e_i$. Write $z = \sum_i z_i e_i \in \mathfrak{h}$ and $x_i = e^{-z_i}$ (so $\prod_i x_i = 1$). Then the

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

left side of (3.4.3) becomes

$$\prod_{1\le i<j\le n}(1-e^{-z_i+z_j}) = x_2^{-1}x_3^{-2}\cdots x_n^{1-n}\prod_{1\le i<j\le n}(x_j-x_i).$$

The right side of (3.4.3) becomes

$$\prod_j x_j^{(n+1-2j)/2}\sum_{\pi\in\mathcal{S}_n}\epsilon(\pi)\prod_i x_{\pi i}^{-(n+1-2i)/2} = x_2^{-1}x_3^{-2}\cdots x_n^{1-n}\sum_{\pi\in\mathcal{S}_n}\epsilon(\pi)\prod_i x_{\pi i}^i.$$

Thus the denominator identity for $\mathfrak{sl}_n(\mathbb{C})$ is simply the formula for the determinant of the Vandermonde matrix

$$\det\begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^n & x_2^n & \cdots & x_n^n \end{pmatrix} = \prod_{1\le i<j\le n}(x_j-x_i). \tag{3.4.4}$$

In the early 1970s Macdonald [**396**] generalised these finite denominator identities to infinite identities, corresponding to the extended Coxeter–Dynkin diagrams. The simplest of his was known classically as the *Jacobi triple product identity*:

$$\prod_{m=1}^{\infty}(1-x^{2m})(1-x^{2m-1}y)(1-x^{2m-1}y^{-1}) = \sum_{n=-\infty}^{\infty}(-1)^n x^{n^2}y^n. \tag{3.4.5a}$$

To Macdonald these were purely combinatorial, but soon Kac, Moody and others reinterpreted his formulae as denominator identities for nontwisted affine algebras, that is substituting $\lambda = 0$ into the Weyl–Kac character formula (3.3.3).

For example, parametrise the Cartan subalgebra of $A_1^{(1)}$ by $z\alpha_1 + z\ell_0 + uC$; then (3.2.5d) says $(m\alpha_1 + n\delta)(z\alpha_1 + \tau\ell_0 + uC) = 2mz - n\tau$. The positive roots of $A_1^{(1)}$ are $\alpha_1 + n\delta$ $(n \ge 0)$, $-\alpha_1 + n\delta$ $(n \ge 1)$ and $n\delta$ $(n \ge 1)$. The Weyl group acts on the Weyl vector ρ by $t_{n\alpha_1}\rho = \rho + 2n\alpha_1 - (2n^2+n)\delta$ and $t_{n\alpha_1}r_{\alpha_1}\rho = \rho + (2n-1)\alpha_1 - (2n^2 - n)\delta$. Thus the $A_1^{(1)}$ denominator identity is

$$\prod_{n=0}^{\infty}(1-rq^n)\prod_{n=1}^{\infty}(1-r^{-1}q^n)\prod_{n=1}^{\infty}(1-q^n) = \sum_{m=-\infty}^{\infty}(-1)^m r^{-m}q^{(m^2+m)/2}, \tag{3.4.5b}$$

where $q = e^{-\tau}$ and $r = e^{-2z}$. Equation (3.4.5a) is recovered from (3.4.5b) by setting $x = \sqrt{q}$ and $y = qr^{-1}$.

Freeman Dyson is a famous quantum physicist, but started his academic life in number theory and still enjoys it as a hobby. Dyson [**166**] found a curious formula for the Ramanujan τ-function, defined by $\sum_{n=1}^{\infty}\tau(n)q^n = \eta(q)^{24} := q\prod_{m=1}^{\infty}(1-q^m)^{24}$:

$$\tau(n) = \sum\frac{\prod_{1\le i<j\le 5}(a_i-a_j)}{1!\,2!\,3!\,4!}, \tag{3.4.6}$$

where the sum is over all 5-tuples a_i with $a_i \equiv i \pmod 5$ obeying $\sum_i a_i = 0$ and $\sum_i a_i^2 = 10n$. Using this, an analogous formula can be found for η^{24}. Dyson knew that similar formulae were also known for η^d for the values $d = 3, 8, 10, 14, 15, 21, 24, 26, 28, 35, 36, \ldots$

What was ironic was that Dyson found (3.4.6) at the same time that Macdonald was finding his own identities. Both were at Princeton then, and would often chat a little when they bumped into each other after dropping off their daughters at school. But they never discussed work. Dyson didn't realise that his strange list of numbers has a simple interpretation: they are precisely the dimensions of the simple Lie algebras! $3 = \dim(\mathrm{A}_1)$, $8 = \dim(A_2)$, $10 = \dim(C_2)$, $14 = \dim(G_2)$, etc. In fact these formulae for η^d are none other than (specialisations of) the Macdonald identities. For example, Dyson's formula is the denominator formula for $A_4{}^{(1)}$ ($24 = \dim(A_4)$). If they had spoken, they would surely have anticipated the affine algebra denominator identity interpretation.

Incidentally, no simple Lie algebra has dimension 26, so the formula for η^{26} can't correspond to any of Macdonald's identities. Its algebraic meaning is still uncertain.

Macdonald certainly didn't close the book on denominator identities. Any algebra with a character formula analogous to (1.5.11) (e.g. Borcherds–Kac–Moody algebras (3.3.4)) will have one. Kac and Wakimoto [**336**] use denominator identities for Lie superalgebras to obtain nice formulae for various generating functions involving sums of squares, sums of triangular numbers (triangular numbers are numbers of the form $\frac{1}{2}k(k+1)$), etc. For instance, the number of ways n can be written as a sum of 16 triangular numbers is

$$\frac{1}{3\cdot 4^3}\sum ab\,(a^2-b^2)^2,$$

where the sum is over all odd positive integers a, b, r, s obeying $ar + bs = 2n + 4$ and $a > b$.

The most important application of denominator identities from our perspective is Borcherds' use of them (Section 7.2.2) in proving the Monstrous Moonshine conjectures. Indeed, this possibility was what motivated his introduction of the Borcherds–Kac–Moody algebras. Other applications are discussed next subsection.

Explicitly writing down denominator identities for Borcherds–Kac–Moody algebras tends to be quite difficult, because their root multiplicities are hard to find. The denominator identity of the Monster Lie algebra $\mathfrak{m}$ is a remarkable identity originally due to Zagier, but discovered independently by Borcherds and others:

$$p^{-1}\prod_{\substack{m>0\\ n\in\mathbb{Z}}}(1-p^m q^n)^{a_{mn}} = J(z) - J(\tau), \tag{3.4.7a}$$

with $p = e^{2\pi \mathrm{i} z}$, where the powers '$a_i$' are the coefficients of the q-expansion of the modular function $J(\tau) = \sum_i a_i q^i$. This yields infinitely many nontrivial polynomial identities in the coefficients a_n – for example, comparing third-degree terms on both sides gives

$$a_4 = \binom{a_1}{2} + a_3. \tag{3.4.7b}$$

In fact, (3.4.7a) is older than $\mathfrak{m}$ and is proved independently (Hecke operators permit a quick proof); turning the logic around, it is used to tell us the root multiplicities of $\mathfrak{m}$. This is its direct use in the proof of the Monstrous Moonshine conjectures.

Unfortunately, the numerator of the Weyl character formula for $L(\lambda)$ rarely has a product formula. However, certain specialisations of the numerator can manifestly equal certain (λ-dependent) specialisations of the denominator, and thus inherit the product expansion of the latter. Consider a simple example: any finite-dimensional A_n-module $L(\lambda)$ has a character satisfying

$$\mathrm{ch}_{L(\lambda)}(t\rho) = x^{(n+1)t(\lambda)/2} \prod_{1\le i<j\le n+1} \frac{x^j y_j - x^i y_i}{y^j - y^i}, \tag{3.4.8a}$$

for any $t \in \mathbb{C}$, where $x = e^t$ and $y_i = \exp[(i - \sum_{j=i}^{n} \lambda_j)t]$. Similar formulae hold for all Kac–Moody algebras [**374**]. In particular, from these we obtain instantly Weyl's dimension formula for finite-dimensional semi-simple Lie algebras:

$$\dim L(\lambda) = \prod_{\alpha>0} \frac{(\alpha|\lambda+\rho)}{(\alpha|\rho)}. \tag{3.4.8b}$$

3.4.3 Automorphic products

In Section 2.4.1 we explain the important notion of *lifting* a modular form $f : \mathbb{H} \to \mathbb{C}$ for a discrete subgroup Γ of $G = \mathrm{SL}_2(\mathbb{R})$. The result is an automorphic function $\phi : G \to \mathbb{C}$ obeying the transformation (2.4.2b).

Borcherds discovered an unexpected way to lift (meromorphic) modular forms for discrete Γ in $\mathrm{SL}_2(\mathbb{R})$ to much larger Lie groups. His starting point was (3.4.7a), where the coefficients of a modular function appear in the exponents of a product expansion. In hindsight, another example of this phenomenon is the product formula(2.2.6b) for η:

$$\eta(\tau) = q^{1/24} \prod_{n=1}^{\infty} (1-q^n)^1, \tag{3.4.9}$$

where the powers '1' are the coefficients of the q-expansion of the modular form $\theta_3(\tau)/2$. Moreover, both (3.4.7a) and (3.4.9) are the denominators of the Monster algebra $\mathfrak{m}$ and the affine algebra $\mathfrak{u}_1{}^{(1)}$ (recall (3.2.12c)). Are these hints of a much more general phenomenon?

Indeed. Borcherds found a far-reaching generalisation of (3.4.7a):

Theorem 3.4.2 [**76**] *Suppose $f(\tau) = \sum_n a_n q^n$ is a meromorphic modular form for $SL_2(\mathbb{Z})$ of weight $-s/2$, holomorphic in $\mathbb{H}$ (so its only possible pole is at the cusp), and with integer coefficients a_n. We require $s = 0, 8, 16, \ldots$; if $s = 0$ we also require that 24 divides a_0. Let $v_0 \in \mathbb{R}^{s+1,1}$ be a generic vector of negative norm. Then there is a unique lattice vector $\rho \in II_{s+1,1} \subset \mathbb{R}^{s+1,1}$ such that*

$$F(v) = e^{-2\pi \mathrm{i}\,\rho\cdot v} \prod_{r\in II_{s+1,1},\; r\cdot v_0>0} (1 - e^{-2\pi \mathrm{i}\, r\cdot v})^{a_{-r\cdot r/2}} \tag{3.4.10}$$

can be analytically extended to a meromorphic modular form on $\mathbb{H}_{s+1,1}$ of weight $a_0/2$ for the group $O_{s+2,2}(\mathbb{Z})^+$.

Since f in Theorem 3.4.2 has nonpositive weight and is holomorphic in $\mathbb{H}$, it will necessarily have poles at the cusps $\mathbb{Q}\cup\{\mathrm{i}\infty\}$ (unless it is constant). The set $II_{s+1,1}$ is the unique even self-dual lattice of signature $(s+1,1)$ (Section 1.2.1). $\mathrm{O}_{s+2,2}(\mathbb{R})$ is the group of $(s+4)\times(s+4)$ matrices A with real entries, which obey $ADA^t=D$ for $D=\mathrm{diag}(1,\ldots,1,-1,-1)$. By a modular form for $\mathrm{O}_{s+2,2}(\mathbb{Z})^+$, we mean the following. First, the imaginary norm vectors in $\mathbb{R}^{s+1,1}$ lie in two disjoint cones; denote by C the cone containing $-v_0$. The analogue of the upper half-plane $\mathbb{H}$ is here the set $\mathbb{H}_{s+1,1}\subset\mathbb{C}^{s+1,1}$ consisting of all vectors v with imaginary part $\mathrm{Im}(v)\in C$. Then

$$F(v+\lambda)=F(v),\qquad \forall\lambda\in II_{s+1,1}, \tag{3.4.11a}$$

$$F(w(v))=\pm F(v),\qquad \forall w\in \mathrm{Aut}(II_{s+1,1})^+, \tag{3.4.11b}$$

$$F\left(\frac{2v}{v\cdot v}\right)=\pm\left(\frac{v\cdot v}{2}\right)^{a_0/2}F(v), \tag{3.4.11c}$$

for appropriate choice of signs, where $\mathrm{Aut}(II_{s+1,1})^+$ are the automorphisms of the lattice $II_{s+1,1}$ that send the cone C to itself. The transformations on $\mathbb{H}_{s+1,1}$ given in (3.4.11) generate a subgroup of $\mathrm{O}_{s+2,2}(\mathbb{Z})$, denoted $\mathrm{O}_{s+2,2}(\mathbb{Z})^+$. Now F can be lifted to the Lie group $\mathrm{O}_{s+2,2}(\mathbb{R})^+$ in the usual way. This lifting of a modular form for a subgroup Γ of $\mathrm{SL}_2(\mathbb{R})$ to automorphic forms for $\mathrm{O}_{s+2,2}(\mathbb{R})^+$ is called a *Borcherds lift*.

Of course (3.4.7a) is recovered from taking $f(\tau)=j(\tau)-744$; then $s=0$, and the real Lie group $\mathrm{O}_{2,2}(\mathbb{R})$ is essentially $\mathrm{SL}_2(\mathbb{R})\times\mathrm{SL}_2(\mathbb{R})$ – that is, they share the same Lie algebra (recall Theorem 1.4.3) – with each $\mathrm{SL}_2(\mathbb{R})$ contributing a copy of $\mathbb{H}$ and $\mathrm{SL}_2(\mathbb{Z})$.

We can recover from F more familiar modular forms by restricting the domain of F to multiples τv of imaginary norm vectors v in $II_{s+1,1}$. For example, we get:

Theorem 3.4.3 [**76**] *Let $f(\tau)=\sum_{n=-\infty}^{\infty}a_nq^n$ be any meromorphic modular form for $\Gamma(4)$, holomorphic in $\mathbb{H}$ but possibly with poles at the cusps, and with integer coefficients a_n. We require $a_n=0$ unless $n\equiv 0,1$ (mod 4). Then for some choice of $h\in\mathbb{Z}/12$,*

$$F(\tau)=q^h\prod_{n=1}^{\infty}(1-q^n)^{a_{n^2}}$$

is a meromorphic modular form of weight a_0, with all poles and zeros at cusps.

For example, (3.4.9) (or rather its square) is recovered by taking $f(\tau)=\theta_3(2\tau)$. Modular forms for SL_2 arise here because $\mathrm{O}_{1,2}(\mathbb{R})$ is essentially $\mathrm{SL}_2(\mathbb{R})$.

In this section we find several examples of product expansions of modular forms, Jacobi forms, etc. coming from the denominators of characters. An exciting development is provided by Gritsenko and Nikulin [**264**], [**265**]. Given any hyperbolic Kac–Moody algebra of rank $n\geq 3$ with certain properties (making them close in spirit to semi-simple Lie algebras), there exists a Borcherds–Kac–Moody algebra of the same rank with identical real roots (hence Weyl group, which will be a subgroup of $\mathrm{O}_{n-1,1}(\mathbb{R})$), but with precisely the imaginary simple roots needed so that its denominator is an automorphic form for $\mathrm{O}_{n,2}(\mathbb{R})$. It is reminiscent of Macdonald's identities: he found he needed to introduce extra factors to get modularity (namely the third product in (3.4.7b)),

https://doi.org/10.1017/9781009401548.004 Published online by Cambridge University Press

and we now interpret those as due to the imaginary roots of the corresponding affine algebra.

Most Borcherds–Kac–Moody algebras are of course not interesting; those that are (e.g. the Monster and fake Monster Lie algebras) have automorphic denominator identities. Thus this provides a systematic construction of what should be interesting Borcherds–Kac–Moody algebras. It is known that there are only finitely many such hyperbolic Kac–Moody algebras, and so this is a finite family of Borcherds–Kac–Moody algebras. Clearly, we should study their representation theory, and compute the characters of their 'interesting' (presumably integrable) modules. In analogy with affine algebras, we may hope that the numerators of those characters will also be automorphic.

Relations of these automorphic forms with mirror symmetry and string theory are beyond this book, but see, for example, [**266**], [**342**], [**275**], [**276**], [**434**]. The review article [**358**] is a good treatment of many of the topics of this subsection.

Question 3.4.1. Let $f(q) = \sum_{n=0} a_n q^n$, with $a_0 = 1$. Verify that, at least formally (i.e. without any regard to convergence), this can be written as $f(q) = \prod_{n=1}^{\infty}(1 - q^n)^{b_n}$ for some unique numbers b_n. If all a_n are integers, then so are all b_n.

Question 3.4.2. Prove (3.4.8a) and the Weyl dimension formula (3.4.8b) for $\mathfrak{sl}_n$.

Question 3.4.3. Express the character χ_λ of any integrable representation λ of $A_1{}^{(1)}$, specialised appropriately, as an infinite product.