primarily algebraic methods. However, they do, and many other representations besides. For example, in **[VoZu]**, all unitary representations with nonvanishing relative $(\mathbf{g}, \mathbf{k})$-cohomology are classified. These are the representations whose multiplicities in $L^2(\Gamma\backslash G)$ determine the Betti numbers of locally symmetric spaces, via Matsushima's formula **[BoWa, HoWa, Mats]**.

4. Other directions and applications. The "applications" of Lie theory are diverse and many; but often it is absurd to speak of "applications" when the role of Lie theory is so basic and pervasive: although $\mathbf{R}^n$ is a Lie group, usually it is used in such a low-tech way that its Lie-theoretic properties are superfluous. But when we combine it with its character group to form the Heisenberg group acting on $L^2(\mathbf{R}^n)$, then its identity as a Lie group becomes more relevant. Similarly with linear algebra: it would be egregious to claim it completely as part of Lie theory, but as I hope was demonstrated in §1, the border beyond which one should definitely consider oneself in Lie-theoretic territory is easy to cross and not so far from the public entrance. There is a similar identity problem on the high-end: to what extent should one include the more exotic algebraic structures dreamed up in physics—Jordan algebras, Kac-Moody algebras, Lie "superalgebras", quantum groups, vertex algebras—in "Lie theory"? Thus a comprehensive survey of "applications of Lie theory" is not simply impossible, it is fruitless. Below I offer instead an eclectic set of examples that I hope hit a few of the high points. For some original and stimulating discussions of many applications of representation theory, not exclusively of Lie groups, see various books and articles of Mackey **[Mack1–3]**.

4.1. *Combinatorics.* Representation theory, even just the finite-dimensional theory of the general linear group, is rife with combinatorial quantities. We illustrate with a few examples.

4.1.1. S-FUNCTIONS. Much of the combinatorics of symmetric functions, developed in the nineteenth century, found natural interpretations in terms of representations of GL_n when that subject began to be understood through the work of Schur **[Schu]** around the turn of the century. The symmetric functions known as S-functions or Schur functions **[Litt, Macd1]** were introduced by Jacobi, but have been named after Schur because of his interpretation of them as the characters of the irreducible "polynomial" representations of GL_n or U_n. There is a famous identity due to Cauchy **[Macd1, Weyl2]** that, when interpreted in terms of representation theory, yields one of the most useful formulations of the fundamental theorems of classical invariant theory.

We will give a brief explanation of Cauchy's identity and its representation-theoretic interpretation. For an extensive discussion of symmetric functions with applications to representation theory, I recommend **[Macd1]**. In this discussion, we will follow the notation of **[Macd1]**, although it differs from our earlier notation.

Let $x_1, x_2, \ldots, x_n$ be indeterminates. We will, when convenient, think of the x_i as coordinates of a point in $\mathbf{C}^n$. Given an n-tuple

$$\mu = (\mu_1, \mu_2, \ldots, \mu_n) \tag{4.1.1.1}$$

of nonnegative integers, we write

$$x^\mu = \prod_j x_j^{\mu_j}. \tag{4.1.1.2}$$

We let the symmetric group S_n permute the variables x_i in the obvious way. This gives rise to a permutation action of S_n on the monomials x^μ:

$$s(x^\mu) = \prod_j x_{s(j)}^{\mu_j}. \tag{4.1.1.3}$$

If an n-tuple $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ of nonnegative integers is arranged in decreasing order, i.e., $\lambda_j \geq \lambda_{j+1}$, we call λ a *partition*. We write

$$\delta = (n-1, n-2, \ldots, 2, 1, 0) \tag{4.1.1.4}$$

for the smallest partition whose entries are strictly decreasing.

For a partition α, define

$$a_\alpha = \sum_{w \in S_n} \varepsilon(w) w(x^\alpha). \tag{4.1.1.5}$$

Here ε is the sign character of S_n. The polynomial a_α is easily checked to be skew-symmetric, i.e.,

$$w(a_\alpha) = \varepsilon(w) a_\alpha.$$

Hence $a_\alpha = 0$ unless the entries α_i of α are strictly decreasing. In this case, we can write $\alpha = \lambda + \delta$, where λ is again a partition. The argument with the Vandermond determinant (cf. §3.5.4) shows that $a_{\lambda+\delta}$ is divisible by a_δ. Set

$$s_\lambda = a_{\lambda+\delta} / a_\delta. \tag{4.1.1.6}$$

Comparison with formula (3.5.4.24), allowing for the differences in notation, shows that if the x_i are thought of as the coordinates of a unitary diagonal matrix, then s_λ is the character of the representation of U_n with highest weight λ. In this context we call the s_λ *Schur-functions* or *S-functions.*

Cauchy's identity says

$$a_\delta(x) a_\delta(y) \left(\prod_{i,j=1}^{n} (1 - x_i y_j) \right)^{-1} = \sum_\lambda a_{\lambda+\delta}(x) a_{\lambda+\delta}(y). \tag{4.1.1.7}$$

The sum is over all partitions λ. This identity arises by evaluating

$$\det((1 - x_i y_j)^{-1})$$

in two different ways (cf. [**Macd1**, pp. 38, 33]). Dividing by $a_\delta(x)a_\delta(y)$ gives us

$$\prod_{i,j=1}^{n}(1-x_iy_j)^{-1}=\sum_{\lambda}s_\lambda(x)s_\lambda(y). \tag{4.1.1.8}$$

The right-hand side of this has a clear representation-theoretic interpretation, at least formally: it is the character of the representation $\sum_\lambda \rho_n^\lambda \otimes \rho_n^\lambda$ of $\mathrm{U}_n \times \mathrm{U}_n$, where ρ_n^λ denotes the representation of U_n with highest weight χ_λ (notation as in equation (3.5.4.2)). This observation challenges us to find a representation-theoretic interpretation of the left-hand side of (4.1.1.8).

The challenge is met by the *Molien formula* [**Bour**, **Moli**, **Stan**]. This is a formula for the trace of an element of $g \in \mathrm{GL}_n$ acting on the space $\mathscr{P}(\mathbf{C}^n)$ of polynomials on $\mathbf{C}^n$. Since $\mathscr{P}(\mathbf{C}^n)$ is infinite dimensional, the notion of trace will require some interpretation. Consider the space $\mathscr{P}^d(\mathbf{C}^n)$ of polynomials homogeneous of degree d. Let us compute the trace of g acting on $\mathscr{P}^d(\mathbf{C}^n)$. Up to conjugation of g, and ignoring the lower-dimensional subvariety consisting of nondiagonalizable g, we may assume g is diagonal, with eigenvalues c_j:

$$g(x_j)=c_jx_j.$$

Then

$$g(x^\mu)=\prod_j g(x_j)^{\mu_j}=\left(\prod_j c_j^{\mu_j}\right)x^\mu=c^\mu x^\mu,$$

where $c=(c_1, c_2, \dots, c_n)$ is the n-tuple of eigenvalues of g. Thus the eigenvalues of g acting on $\mathscr{P}^d(\mathbf{C}^n)$ are $\{c^\mu\}$, where $|\mu|=\sum\mu_j=d$. Thus

$$\operatorname{trace} g_{|\mathscr{P}^d(\mathbf{C}^n)}=\sum_{|\mu|=d}c^\mu.$$

If now we formally sum over all d we get

(4.1.1.9)

$$\begin{aligned}\operatorname{trace} g_{|\mathscr{P}(\mathbf{C}^n)}&=\sum_\mu c^\mu=\sum_{\mu_j\geq 0}c_1^{\mu_1}c_2^{\mu_2}\cdots c_n^{\mu_n}\\&=\left(\sum_{\mu_1\geq 0}c_1^{\mu_1}\right)\left(\sum_{\mu_2\geq 0}c_2^{\mu_2}\right)\cdots\left(\sum_{\mu_n\geq 0}c_n^{\mu_n}\right)\\&=\prod_{j=1}^{n}\frac{1}{(1-c_j)}=\left(\prod_{j=1}^{n}1-c_j\right)^{-1}=(\det(1-g))^{-1}.\end{aligned}$$

Here 1 indicates the $n\times n$ identity matrix. Observe that if all the eigenvalues c_j of g have absolute value less than 1, then the infinite series in (4.1.1.9) converges as $d\to\infty$. Thus equation (4.1.1.9) can be regarded as an equality of analytic functions on the appropriate open subset of $\mathrm{GL}_n(\mathbf{C})$; or since the right-hand side of the equation is rational, it can be thought of as an interpretation of the left-hand side as a rational function.

If $G \subseteq \mathrm{GL}_n(\mathbf{C})$ is a subgroup, then G may not contain any elements whose eigenvalues are all of absolute value less than 1 (i.e., $G = \mathrm{SL}_n(\mathbf{C})$, etc). However, if we augment G by the group $\mathbf{C}^x$ of scalar operators, this larger group will clearly have such elements. By this device, formula (4.1.1.19) makes sense for elements of any subgroup of $\mathrm{GL}_n(\mathbf{C})$. If we wish to emphasize the role of the scalar, we can write

$$\text{trace } tg_{|\mathscr{P}(\mathbf{C}^n)} = (\det(1 - tg))^{-1}. \tag{4.1.1.10}$$

This formula is valid for all $g \in \mathrm{GL}_n(\mathbf{C})$, and all sufficiently small t (depending on g) in $\mathbf{C}$. It can also be regarded as an identity in formal power series in t. Thus if $\pi : G \to \mathrm{GL}_n(\mathbf{C})$ is a representation of $\mathbf{C}^n$, we can write

$$\mathrm{trace}(t\pi(g)_{|\mathscr{P}(\mathbf{C}^n)}) = (\det(1 - t\pi(g)))^{-1}. \tag{4.1.1.11a}$$

We can regard equation (4.1.1.11a) as the correct version of the equation

$$\mathrm{trace}(\pi(g)_{|\mathscr{P}(\mathbf{C}^n)}) = (\det(1 - \pi(g)))^{-1} \tag{4.1.1.11b}$$

which will make sense as long as the set of g in G such that $\pi(g)$ has 1 as an eigenvalue is nowhere dense in G.

Given this background, we see our challenge is to find a representation of $\mathrm{U}_n \times \mathrm{U}_n$ such that the left-hand side of equation (4.1.1.8) is the right-hand side of equation (4.1.1.11b) (for g a product of diagonal matrices). We do not have far to seek. Consider the action π of $\mathrm{U}_n \times \mathrm{U}_n$ on $M_n(\mathbf{C})$, the $n \times n$ matrices, by left and right multiplication:

$$\pi(g_1, g_2)(T) = g_1 T g_2^t, \qquad g_i \in \mathrm{U}_n,\ T \in M_n. \tag{4.1.1.12}$$

Here g_2^t is the transpose of g_2; we use g_2^t instead of the more standard g_2^{-1} in order to make things come out symmetrically in g_1 and g_2. It is trivial to check that if g_1, g_2 are diagonal matrices with eigenvalues $\{x_j\}_{j=1}^n$ and $\{y_k\}_{k=1}^n$ respectively, then the right-hand side of (4.1.1.11b) is exactly the left-hand side of (4.1.1.8). Thus (4.1.1.8) is revealed as an expansion of $\mathrm{trace}(\pi(g_1, g_2)_{|\mathscr{P}(M_n)})$ into a sum of characters of $\mathrm{U}_n \times \mathrm{U}_n$. Because representations are determined by their characters, we deduce the following corollary of the combination of the Cauchy identity and Molien's formula.

THEOREM 4.1.1.13 (Fundamental theorem of invariant theory, polynomial duality version). *Under the action π (cf. (4.1.1.12)) of $\mathrm{U}_n \times \mathrm{U}_n$ on $M_n(\mathbf{C})$, the decomposition of $\mathscr{P}(M_n(\mathbf{C}))$ into irreducible representations is described by*

$$\mathscr{P}(M_n(\mathbf{C})) \simeq \sum_{\lambda} \rho_n^{\lambda} \otimes \rho_n^{\lambda}.$$

Here λ runs over all partitions (cf. (4.1.1.1)–(4.1.1.4)). The subspace $\mathscr{P}^d(M_n(\mathbf{C}))$ is the sum over all λ with $|\lambda| = \sum_{j=1}^n \lambda_j = d$.

REMARKS. (a) A significant feature of this result is that, in the decomposition of $\mathscr{P}(M_n(\mathbf{C}))$, a given representation ρ_n^λ of the first factor of $\mathrm{U}_n \times \mathrm{U}_n$ is combined with exactly one representation (which again happens to be ρ_n^λ) of the second factor. In other words, the symmetry type under left multiplication of a polynomial on $M_n(\mathbf{C})$ determines its symmetry type under right multiplication. This fact has many repercussions in invariant theory. In particular, it directly implies various reciprocity laws (cf. **[Howe8]**).

(b) In the discussion above, we have used a combinatorial formula to make a representation-theoretic conclusion. However, the flow could be reversed: we can prove Theorem 4.1.1.13 directly from general principles of representation theory. Computing the character of the action via Molien's formula (4.1.1.12) then would yield Cauchy's identity in the form (4.1.1.8). The main observation of the representation-theoretic approach is that Theorem 4.1.1.13 is essentially an example of the Peter-Weyl Theorem (cf. §3.5.4). To see this, we make the following observations.

(i) Polynomial functions on $M_n(\mathbf{C})$ are determined by their restriction to U_n. This is because polynomials are in particular holomorphic functions on $M_n(\mathbf{C})$, and U_n is a "real form" of $\mathrm{GL}_n(\mathbf{C})$, in the sense that its Lie algebra $\mathbf{u}_n$ is a real subspace of $\mathbf{gl}_n(\mathbf{C}) \simeq M_n(\mathbf{C})$ such that $\mathbf{gl}_n(\mathbf{C}) = \mathbf{u}_n \oplus i\mathbf{u}_n$ (cf. §3.5.5). It follows that if p is a polynomial on $M_n(\mathbf{C})$, then the Taylor series of p at 1, the identity matrix, is determined by the Taylor series of $p_{|\mathrm{U}_n}$, the restriction of p to U_n. Since p is in turn determined by its Taylor series at 1, the observation follows.

(ii) The restriction of $\mathscr{P}(M_n(\mathbf{C}))$ to U_n will yield a space of functions which is invariant under left and right multiplications on U_n. Hence the Peter-Weyl Theorem (cf. Theorem 3.5.4.23) implies that $\mathscr{P}(M_n(\mathbf{C}))$ is a direct sum of modules of the form $\rho_n^\lambda \otimes (\rho_n^\lambda)^*$, where $(\rho_n^\lambda)^*$ is the contragredient of ρ_n^λ. The automorphism $g \to (g^t)^{-1}$, used to create the action π of equation (4.1.1.12), transforms $(\rho_n^\lambda)^*$ to ρ_n^λ.

(iii) From (i) and (ii) we conclude that $\mathscr{P}(M_n(\mathbf{C}))$ is a sum of $(\mathrm{U}_n \times \mathrm{U}_n)$-modules of the form $\rho_n^\lambda \otimes \rho_n^\lambda$. It remains to determine which λ can occur. Since all the weights of the diagonal torus in $\mathrm{U}_n \times \mathrm{U}_n$ are polynomial weights, in the sense that they involve only nonnegative powers of the diagonal entries, it follows that the highest weights λ must be restricted to be partitions, i.e., must also have all entries nonnegative. Conversely, it is possible to explicitly exhibit $\mathrm{U}_n \times \mathrm{U}_n$ highest weight vectors of weight λ, for any partition λ. Let $\{z_{ij} : 1 \leq i, j \leq n\}$ denote the standard matrix-entry coordinates on $M_n(\mathbf{C})$. Set

$$\gamma_k = \det \begin{vmatrix} z_{11} & z_{12} & \cdots & z_{1k} \\ z_{21} & & & \vdots \\ \vdots & \ddots & & \\ z_{k1} & & \cdots & z_{kk} \end{vmatrix}$$

for $1 \le k \le n$, and, for a partition λ,

$$\gamma^{\lambda} = \prod_{k \ge 1} (\gamma_k)^{\lambda_k - \lambda_{k-1}} \tag{4.1.1.14}$$

(where we agree that $\lambda_{n+1} = 0$). Then γ^{λ} is a $U_n \times U_n$ highest weight vector of weight λ, showing that $\rho_n^{\lambda} \otimes \rho_n^{\lambda}$ does occur as a $(U_n \times U_n)$-submodule of $\mathscr{P}(M_n(\mathbf{C}))$.

(c) The above argument is perhaps an expensive way to establish Cauchy's identity, but it does illustrate in a simple case the potential combinatorial import of the seemingly bland general structure theorems of representation theory. The combinatorial content, of course, is supplied by the structure of Lie groups, in this case U_n. Also, it is important to understand that Theorem 4.1.1.13 is a robust result, not dependent on clever manipulations of power series.

(d) For further details on how Theorem 4.1.1.13 leads to invariant theory, we refer to [**Howe1, 8, 9**]. For more examples of identities that arise in similar fashion, see [**Proc2, Litt, Tera**].

4.1.2. Multiplicities. A basic problem in the representation theory of U_n (or any other compact Lie group), is understanding how an irreducible representation of U_n decomposes under restriction to the Cartan subgroup (diagonal torus) A. This problem is of interest in itself; additionally, it has many ramifications, some quite surprising. Our goal in this section is to explain some of these ramifications.

Let ρ be a representation of U_n on a vector space V. We can decompose V into eigenspaces for the diagonal torus A:

$$V = \sum_{\chi \in \widehat{A}} V_{\chi}, \tag{4.1.2.1}$$

where

$$\rho(a)(v) = \chi(a)v, \qquad a \in A,\ v \in V_{\chi}.$$

The space V_{χ} is called the *χ-weight space*. The collection of $\{V_{\chi}\}$ are called the *A weight spaces*, or just the weight spaces. The dimension $\dim V_{\chi}$ is called the *multiplicity* of χ in ρ. We are particularly interested in the multiplicity of χ when $\rho = \rho_n^{\lambda}$ is an irreducible representation of U_n. We denote the multiplicity of χ in ρ_n^{λ} by

$$m(\lambda, \chi). \tag{4.1.2.2}$$

The Weyl group S_n normalizes A inside U_n, and so it will act on the decomposition (4.1.2.1), permuting the V_{χ}. Thus $\dim V_{\chi} = \dim V_{w(\chi)}$ for $\chi \in \widehat{A}$. Hence in computing the $m(\lambda, \chi)$, one can restrict to $\chi \in \widehat{A}^{+}$ (cf. definition (3.5.4.15)). In other words, if we take $\chi = \chi_{\alpha}$, where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ is in $\mathbf{Z}^n$ (cf. equation (3.5.4.2)), we can assume that the entries α_j of α are decreasing. If all the α_j are nonnegative, then α will be a partition. If $\chi = \chi_\alpha$, for α a partition, we will write

$$m(\lambda, \chi_\alpha) = K_{\lambda\alpha}. \tag{4.1.2.3}$$

In some sense, the problem of computing the $m(\lambda, \chi)$ is solved. There is a general formula due to Kostant [**Hump**, **Jaco1**, **Kost5**] for the multiplicity of a weight in an irreducible representation that applies to any semisimple Lie group. It amounts to a suitable rewriting of the Weyl character formula. There is also a recursive method for computing multiplicities given by Freudenthal [**Jaco1**], which has been made much more efficient by Moody and Patera [**MoPa**, **BrMP**]. For U_n, there is a combinatorial description of the multiplicities that goes back to Kostka [**Kstk**], and has its representation-theoretic interpretation in the Gelfand-Cetlin basis [**BiLo2**, **Zhel1**, **Cher1**]. There is an analog of this latter for more general groups in the standard monomial theory of Seshadri et al. [**Sesh**, **LaSe1**, **2**, **LaMS**]. Thus, in some sense, the determination of the $m(\lambda, \chi)$ is a very well-solved problem. On the other hand, there are problems for which these standard solutions are of little help, so there is still considerable room for a deeper understanding of the $m(\lambda, \chi)$.

Our purpose here is not to compute the $m(\lambda, \chi)$, but to show how these numbers, which arise in this context, reverberate through representation theory, echoing from topic to topic until they reach contexts semingly quite removed from each other.

The first reflected image of the $m(\lambda, \chi)$ is in the theory of symmetric functions. This is formed by the S-functions as the characters of the irreducible representations of U_n. The S-functions are defined (cf. equation (4.1.1.6)) in a relatively sophisticated and indirect way via a quotient. The most simple minded symmetric function to associate to a partition is the symmetrization of a monomial:

$$m_\alpha = \#(W_\alpha)^{-1} \sum_{w \in S_\alpha} w(x^\alpha). \tag{4.1.2.4}$$

Here α is a partition and $W_\alpha \subseteq S_n$ is the stabilizer of α. It is obvious that the m_α form a basis for the space of symmetric functions. In order to understand the S-functions s_λ, we might try to express s_λ as a linear combination of the m_α. It is not hard to see that the desired expression is

$$s_\lambda = \sum K_{\lambda\alpha} m_\alpha. \tag{4.1.2.5}$$

Indeed, it is clear that the contribution of $\sum_{w \in S_n} V_{\chi_{w(\alpha)}}$ to the trace of $a \in A$ with entries $(z_1, \dots, z_n)$ is just $(\dim V_{\chi_\alpha}) m_\alpha(z)$; equation (4.1.2.5) is immediate from this and definition (4.1.2.3). It is in this role, as coefficients for expressing the s_λ in terms of the m_α, that the $K_{\lambda\alpha}$ first appeared. As such, they were studied by Kostka [**Kstk**], hence are known as *Kostka coefficients*.

Perhaps the most simple-minded occurrence of the Kostka coefficients outside their initial role as weight multiplicities is as K-multiplicities for principal series representations of $\mathrm{GL}_n(\mathbf{C})$. We refer to §3.6.1 for the context of these remarks. As explained in §3.6.5, the study of representations of a semisimple group G by means of their restrictions to the maximal compact subgroup K of G is a basic technique. Of course a representation of K is determined by the multiplicities with which the various irreducible representations of K occur. Since the principal series are such an important class of representations of G, we are curious to know the multiplicities of K-types (i.e., irreducible representations of K) in principal series representations of G. Because of the Iwasawa decomposition $G = KP$ for any parabolic subgroup P (cf. §A.2.3.5), the multiplicity problem for principal series reduces to a problem in the representation theory of K. Indeed, we can see directly from the Iwasawa decomposition and the definition (3.6.1.6), that

$$\text{P.S.}(\sigma, \psi)_{|K} \simeq \mathrm{ind}_{K\cap M}^{K}(\sigma_{|K\cap M}). \tag{4.1.2.6}$$

For the minimal parabolic P_0, we have $K \cap M_0 = M_0$; also σ is an irreducible representation of M_0, so formula (4.2.1.6) simplifies to

$$\text{P.S.}(\sigma, \psi)_{|K} \simeq \mathrm{ind}_{M_0}^{K} \sigma. \tag{4.1.2.7}$$

Thus the multiplicities of K-types for these principal series are governed by Frobenius reciprocity (cf. **[Gaal, HeRo, Knap2]**, etc.):

$$m(\tau, \text{P.S.}(\sigma, \psi)_{|K}) = m(\sigma, \tau_{|M}), \qquad \tau \in \widehat{K}, \sigma \in \widehat{M_0}. \tag{4.1.2.8}$$

Here the left-hand side of the equation indicates the multiplicity of the irreducible representation τ of K in the principal series $\text{P.S.}(\sigma, \psi)$, $\sigma \in \widehat{M_0}$, and the right-hand side denotes the multiplicity of σ in the restriction of τ to M_0.

Consider the example of $G = \mathrm{GL}_n(\mathbf{C})$. Then $K = \mathrm{U}_n$, and $M_0 = A$, the diagonal torus. Hence σ is just a character of A, and the multiplicity of τ in $\text{P.S.}(\sigma, \tau)$ is just the multiplicity of σ in τ, i.e., is a Kostka coefficient.

The next appearance of the $K_{\lambda\alpha}$ is as decomposition numbers for tensor products. This is an example of the reciprocity law associated to Theorem 4.1.1.13. Let $\mathscr{P}^a(\mathbf{C}^n)$ be the space of polynomials of degree a on $\mathbf{C}^n$. It defines an irreducible representation of U_n, of a relatively simple and comprehensible sort. In particular, all the weight spaces have dimension 1. One might hope to understand more complicated representations of U_n in terms of the $\mathscr{P}^a(\mathbf{C}^n)$. To that end, consider a tensor product

$$\mathscr{P}^{\alpha_1}(\mathbf{C}^n) \otimes \mathscr{P}^{\alpha_2}(\mathbf{C}^n) \otimes \cdots \otimes \mathscr{P}^{\alpha_n}(\mathbf{C}^n).$$

This defines a representation of U_n, probably not irreducible. The reason for considering a tensor product involving n factors is that this is what is needed to obtain an arbitrary polynomial representation of U_n as a constituent. If we want to consider fewer than n factors, we can just let some of the α_j

equal zero. Without loss of generality, we can arrange for the α_j to decrease in j, so that $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ is a partition. Let us then consider the decomposition

$$\mathscr{P}^{\alpha_1}(\mathbf{C}^n) \otimes \mathscr{P}^{\alpha_2}(\mathbf{C}^n) \otimes \cdots \otimes \mathscr{P}^{\alpha_n}(\mathbf{C}^n) \simeq \sum n(\alpha, \lambda)\rho_n^\lambda \tag{4.1.2.9}$$

of the tensor product of the $\mathscr{P}^{\alpha_j}(\mathbf{C}^n)$. It turns out that

$$n(\alpha, \lambda) = K_{\lambda\alpha}. \tag{4.1.2.10}$$

This can be deduced as an example of the reciprocity laws which follow from Theorem 4.1.1.13. One can identify the tensor product on the left side of (4.1.2.9) as a subspace of $\mathscr{P}(M_n(\mathbf{C}))$ by regarding the variables of the kth factor $\mathscr{P}^{\alpha_k}(\mathbf{C}^n)$ to be the entries $\{z_{jk} : 1 \leq j \leq n\}$ of the kth column of the matrix $M = \{z_{jk} : 1 \leq j, k \leq n\}$. Since the homogeneous pieces $\mathscr{P}^a(\mathbf{C}^n) \subseteq \mathscr{P}(\mathbf{C}^n)$ are precisely the eigenspaces for the center of U_n, it is not hard to convince oneself that the tensor product in (4.1.2.9) is precisely the χ_α-eigenspace for the diagonal torus of U_n acting on $M_n(\mathbf{C})$ by multiplication on the right, i.e., the torus in the second factor in the action π of formula (4.1.1.12). Combining this observation with Theorem 4.1.1.13, we see that the ρ_n^λ-isotypic component of the tensor product in (4.1.2.9) is equal to the χ_α-eigenspace of A in the second factor of $\mathrm{U}_n \times \mathrm{U}_n$, in the representation $\rho_n^\lambda \otimes \rho_n^\lambda$. It follows that the multiplicity of ρ_n^λ in the tensor product is the multiplicity of χ_α in ρ_n^λ, whence formula (4.1.2.10). In fact, the symmetric function equivalent of (4.1.2.10) was already known to Kostka.

The $K_{\lambda\alpha}$ also have an interpretation in terms of the representation theory of the symmetric group. This results from Schur duality **[Howe8; Weyl2**, Chapter 4] in the same way that (4.1.2.10) followed from the $(\mathrm{GL}_n, \mathrm{GL}_n)$-Duality Theorem 4.1.1.13.

We recall just enough representation theory of S_n to state the result. The irreducible representations of S_n can be parametrized in a reasonable way by partitions of size n (cf. **[Weyl2, Jaco2, Litt]**, etc.). If $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a partition of size n, i.e., with $\sum \lambda_i = n$, let σ^λ be the associated irreducible representation of S_n.

To a partition α of size n we may associate a (conjugacy class of) subgroup of S_n. Namely, we partition the set $\{1, 2, 3, \ldots, n\}$ into subsets of size α_j, and consider the subgroup which preserves each of the chosen subsets. We denote this subgroup S_α. Form the induced representation $\mathrm{ind}_{S_\alpha}^{S_n} 1$. The decomposition of this representation into irreducible constituents is described by the Kostka coefficients:

$$\mathrm{ind}_{S_\alpha}^{S_n} \simeq \sum K_{\lambda\alpha}\sigma^\lambda \tag{4.1.2.11}$$

REMARKS. There is a natural partial order on partitions obtained from thinking of them as elements of the positive Weyl chamber for U_n. We say

$\lambda \geq \mu$ provided $\sum_{j=1}^{l} \lambda_j \geq \sum_{j=1}^{l} \mu_j$ for all l, $1 \leq l \leq n$. The representation σ^λ is characterized by the fact that σ^λ occurs in $\mathrm{ind}_{S_\lambda}^{S_n} 1$ (with multiplicity 1), but is not contained in $\mathrm{ind}_{S_\mu}^{S_n} 1$ unless $\lambda \geq \mu$. This follows from equation (4.1.2.11), but is usually proved in establishing the existence of the σ^λ.

Finally, we indicate a further appearance of the $K_{\lambda\alpha}$ as decomposition numbers, this time for GL_n over *finite* fields. Let F_q be the finite field with q elements. Let $\mathrm{GL}_n(F_q) = G$ be the group of invertible matrices with coefficients in F_q. Let B be the Borel subgroup of upper triangular matrices in G. We want to consider the induced representation

$$\tau_B = \mathrm{ind}_B^G\ 1,$$

i.e., the natural action of G on $L^2(G/B)$.

Let $P \supseteq B$ be a parabolic subgroup of G containing B. Then P is a group of block upper triangular matrices for some choice of block sizes $\beta_1, \beta_2, \ldots, \beta_k$ with $\sum \beta_j = n$. We can consider the representations

$$\tau_P = \mathrm{ind}_P^G\ 1$$

which will be subrepresentations of τ_B. For a given parabolic P, with block sizes $\beta_1, \beta_2, \ldots, \beta_k$ consider another parabolic P', whose block sizes are $\beta'_1, \beta'_2, \ldots, \beta'_k$, where $\beta'_j = \beta_{s(j)}$ are a permutation of the block sizes of P. We call two parabolics related in this way *associate parabolics.* Although distinct associate parabolics are not conjugate in G, nevertheless we have $\tau_P \simeq \tau_{P'}$. Hence, the representation τ_P depends only on the partition of n defined by the block sizes β_j. Thus, without loss of generality, we may assume the β_j are arranged in order of size: $\beta_j \geq \beta_{j+1}$. Supplementing the β_j with $n-k$ zeros, we can attach to P a partition $\beta = (\beta_1, \beta_2, \ldots, \beta_k, 0, 0, \ldots, 0)$, and we have a well-defined representation $\tau_\beta = \tau_P$ associated to β.

From the Bruhat decomposition for GL_n (cf. §1.2), we can conclude that the intertwining number (i.e., $\dim \mathrm{Hom}_{\mathrm{GL}_n}$; cf. §A.1.4) of τ_β and τ_α is equal to the intertwining number of the representations $\mathrm{ind}_{S_\beta}^{S_n} 1$ and $\mathrm{ind}_{S_\alpha}^{S_n} 1$ of S_n (cf. **[HoMy1, Lusz1]**). From this and the fact that the representations $\mathrm{ind}_{S_\beta}^{S_n} 1$ form a basis for the representation ring of S_n (cf. the Remark following equations (4.1.2.11)), we can conclude that the constituents of τ_B can be labeled by partitions in such a way that

$$\tau_\alpha \simeq \sum K_{\lambda\alpha} \tau_\lambda^0, \tag{4.1.2.12}$$

where τ_λ^0 is the irreducible representation labeled by the partition λ. We see that the τ_λ^0 can be characterized in a fashion similar to the irreducible representations σ^λ of S_n: τ_λ^0 does not occur in τ_α unless $\alpha \leq \lambda$, and it occurs in τ_λ with multiplicity 1.

For a fuller account of the above facts, we refer to [**HoLe**, **Iwah1**, **HoMy1**, **Lusz1**]. The argument just sketched applies only to GL_n : the Bruhat Decomposition is of course available for general reductive groups; however, the representation ring is spanned by the representations $\mathrm{ind}_{W_P}^{W} 1$, for Weyl groups W_P of parabolic subgroups, only for Weyl groups of type A_n .

4.1.3: CURRENT EVENTS. (i) *Symmetric functions.* The characters of U_n are not the only symmetric functions which arise naturally in representation theory of Lie groups. Other examples are functions on $\mathrm{GL}_n(\mathbf{R})$ which are invariant under both right and left translation by O_n . Such functions which are polynomials and eigenfunctions for the center of the universal enveloping algebra for $\mathrm{GL}_n(\mathbf{R})$ are called "zonal spherical polynomials" for $\mathrm{GL}_n(\mathbf{R})$. They are matrix coefficients for O_n-invariant vectors in the finite dimensional irreducible representations of $\mathrm{GL}_n(\mathbf{R})$ (which are in natural correspondence with the representations of U_n). They are used particularly in probability [**Jame**, **Meht**] and also elsewhere [**Dyso1**, **GrRi1**, **2**]. There is an analogous family of polynomials, also parametrized by partitions, associated to $\mathrm{GL}_n(\mathbf{H})$, the group of nonsingular $n \times n$ matrices with entries in $\mathbf{H}$, the quaternions.

In addition to these, other symmetric functions arise naturally from p-adic groups [**Macd1**]. Recently I. G. Macdonald has defined a two-parameter family of bases of the symmetric functions, which include all the above functions for appropriate values of the parameters [**Macd4**]. These are currently under active study (cf. [**Cher2**, **Dunk**, **Heck1**, **2**, **HeOp**, **Opda1–3**], etc.).

These developments grow out of earlier work of Macdonald [**Macd3**], which established a class of identities involving the Dedekind η-function:

$$\eta = q^{1/24} \prod_{l=1}^{\infty} (1 - q^l). \tag{4.1.3.1}$$

Except for the factor $q^{1/24}$, the function η is the reciprocal of the generating function for the partition function [**Andr1**]; thus it is an object of deep combinatorial interest. Macdonald established one identity for each "affine root system" [**Bour**, **Morr**, **Hump**, **Hill**]. The identity for the system $\widetilde{A}_1$ is the Jacobi triple product identity (cf. [**Andr2**], etc.); for the system $\widetilde{BC_1}$ it is the "quintuple product identity," which is usually attributed to Watson, but which can be found in Fricke-Klein. (The identities for the classical root systems were found by Dyson [**Dyso1**, **2**], who, however, did not see the connection with root systems.) Macdonald saw his identities as analogs for affine root systems of the Weyl denominator identity (cf. [**Jaco1**, §VIII.3; **Hump**, §24.3], etc.) for finite root systems. (For root systems of type A_n , this is the formula for the Vandermond determinant, cf. formula (3.5.4.20).) Somewhat after his work, a full analog of the Weyl character formula was established for appropriate representations of "affine Lie algebras" and even more general Kac-Moody algebras [**Kac4**, **GaLe**]. In a more recent reconsideration of his identities, Macdonald drew attention to certain finite truncations of the infinite products involved and formulated some conjectures regarding their evaluation. Macdonald's conjectures were also extended by Morris [**Morr**]. In a flurry of activity by numerous authors, these have recently been established for most affine root systems [**GaGo**, **Gust**, **Habs**, **Stem**, **Zeil1**, **2**].

We also mention a closely related, but rather different line of work pursued by Gustafson and Milne [**GuMi1–3**], who formulate a notion of "well-poised hypergeometric series." These ideas are inspired by the efforts of Biedenharn, with Louck and others [**BiLo2**], to understand tensor products of representations in an explicit way.

(ii) *Kazhdan-Lusztig polynomials.* The rather different incarnations of the Kostka coefficients, involving both infinite-dimensional representations of Lie groups, and representations of finite reductive groups, provides an easily grasped example of the intimate connections that exist between these superficially different topics. Probably the most striking such example is provided by the Kazhdan-Lusztig polynomials [**KaLu1**, **Shi**]. These polynomials are defined in quite a technical way, in connection with what might appear to be a minor issue in the structure theory of "Hecke algebras for G/B"; but they turn out to have some extraordinary connections with phenomena in topology, algebraic geometry, and representation theory of Lie groups [**Shi**, **Lu**, **BoBM**].

We review briefly the construction/definition of the Kazhdan-Lusztig polynomials. Let Σ be a finite root system, W the associated Weyl group, and R a generating set of fundamental reflections. Define the Hecke algebra $\mathscr{H}_W$ associated to W to be the algebra over the field $\mathbf{C}(q^{1/2})$ of rational functions in an indeterminate $q^{1/2}$ generated by elements T_s, $s \in R$, subject to relations

(4.1.3.2a) $$T_s^2 = (q-1)T_s + q,$$

(4.1.3.2b) $$T_s^{\beta_{rs}}(T_rT_s)^{\alpha_{rs}} = T_r^{\beta_{rs}}(T_sT_r)^{\alpha_{rs}}, \qquad r \neq s;\, r, s \in R,$$

where $m_{rs} = 2\alpha_{rs} + \beta_{rs}$, $0 \leq \beta_{rs} \leq 1$, is the order of the element rs in W.

Some readers may wonder why we use $\mathbf{C}(q^{1/2})$ as coefficients, when only q, not $q^{1/2}$, is involved in the defining relations (4.1.3.2). These readers should ask Lusztig.

If we specialize q to be the power of a prime, then $\mathscr{H}_W$ has a direct interpretation as the algebra of intertwining operators for the induced representation $\mathrm{ind}_B^G\, 1$, where G is the Chevalley group, over the finite field F_q, associated to the root system R, and $B \subseteq G$ is a Borel subgroup [**Spri**, **Crtr**]. If $q = 1$, then relations (4.1.3.2) just reduce to the defining relations for W, so we get the group algebra of W. Thus $\mathscr{H}_W$ may be thought of as a one-parameter family of algebras containing $\mathbf{C}(W)$ and all the intertwining algebras for $\mathrm{ind}_{B(F_q)}^{G(F_q)}\, 1$ as F_q varies over all finite fields; in other words, we may regard the intertwining algebras as "deformations" of $\mathbf{C}(W)$. Using this observation it is fairly easy to show **[Iwah2]** that all the intertwining algebras are isomorphic to $\mathbf{C}(W)$. This fact accounts, at least in a philosophical way, for phenomena such as the persistence of the Kostka coefficients across characteristics.

Consider an element w in W. Factor

$$w = s_1 s_2 \cdots s_k, \qquad s_j \in R. \tag{4.1.3.3}$$

The minimum number k of factors in equation (4.1.3.3) is called the *length* of w and written $l(w)$. Given a factorization (4.1.3.3) of w, with $k = l(w)$, define

$$T_w = T_{s_1} T_{s_2} \cdots T_{s_k}. \tag{4.1.3.4}$$

The relation (4.1.3.2)(b) guarantees that T_w is well defined, i.e., independent of the minimal length factorization (4.1.3.3) of w. Then the relations (4.1.3.2)(a) and (b) together imply that the T_w, $w \in W$, define a basis of $\mathscr{H}_W$.

We recall **[Hilr, Dixm1]** that there is defined on W a partial order, the *Bruhat order.* We say $u \leq w$ if there is some minimal expression (4.1.3.3) for W from which we may obtain u by simply deleting some of the s_j. In geometric terms, if G is the (say, complex) semisimple group attached to the root system Σ and $B \subseteq G$ is the Borel subgroup for which R is the set of fundamental generators of W, then $u \leq w$ if and only if BuB is contained in the (Zariski) closure of BwB.

Following Kazhdan and Lusztig **[KaLu1]** we define an involution $a \to \overline{a}$ of the algebra $\mathscr{H}_W$ as follows:

$$\overline{q} = q^{-1} \quad (\text{i.e., } (q^{1/2})^{-} = q^{-1/2}), \qquad \overline{T}_w = (T_w)^{-1} \tag{4.1.3.5}$$

It is easy to check that definitions (4.1.3.5) preserve the defining relations (4.1.3.2) of $\mathscr{H}_W$, so that $a \to \overline{a}$ extends uniquely to an automorphism of $\mathscr{H}_W$ (as an algebra over $\mathbf{C}$).

THEOREM 4.1.3.6 **[KaLu1]**. *For each pair (y, w) of elements of W, $y \leq w$, there is a unique polynomial $P_{y,w}(q)$ of degree at most $\frac{1}{2}(l(w)-l(y)-1)$, such that $P_{ww} = 1$ for each $w \in W$, and the element*

$$C_w = \sum_{y \leq w} (-1)^{l(w)+l(y)} q^{l(w)/2-l(y)} \overline{P}_{y,w} T_w$$

satisfies $C_w = \overline{C}_w$.

This is proved by induction on w. For example,

$$C_s = q^{-1/2} T_s - q^{1/2}, \qquad s \in R.$$

The $P_{y,w}$ are the *Kazhdan-Lusztig polynomials.* They have the following rather amazing connections with geometry and representation theory:

(4.1.3.7)(a) The $P_{y,w}$ are the Poincaré polynomials for the local intersection cohomology (see [**GoMP1**, **2**, **Kirw3**] for a description of this) at y in the Schubert variety $((BwB)/B)^-$ in the flag variety G/B. Here G, B are, as above, the complex Chevalley group and Borel subgroup associated to Σ and R. See [**Crtr**, **Spri**].

(b) The values of $P_{y,w}$ at $q = 1$ describe how the Verma modules (cf. §3.5.3) of fixed infinitesimal characters break up into irreducible highest weight modules. This transfers to a description of the composition series of principal series for complex Lie groups, at least when the infinitesimal character is the infinitesimal character of a finite-dimensional representation. This was conjectured in [**KaLu**], and proved by Bernstein-Beilinson [**BeBe**] and Brylinski-Kashiwara [**BrKa**]. The description of composition series was extended to all semisimple Lie groups by Vogan [**Voga7**].

The Kazhdan-Lusztig polynomials can be used to express other quantities of interest in representation theory, notably the structure of primitive ideals in the universal enveloping algebra of a semisimple Lie algebra [**BeBe**, **BrKa**, **Shi**]. They can also be used to express the Kostka coefficients [**Lusz3**; **Shi**, §2.7]. Unfortunately, the $P_{y,w}$ are themselves rather difficult to compute. Nevertheless, the defining conditions of Theorem 4.1.3.6 define the $P_{y,w}$ recursively, so they can in principle be computed mechanically. This "in principle" caveat applies to many Lie-theoretic quantities one would like to know. It is a challenge to devise more effective means of computation in particular situations of interest.

4.2. *Automorphic forms.* The theory of automorphic forms is a major customer and source of problems for representation theory, to the extent that it is sometimes difficult to draw a boundary between the two fields. This state of affairs is the culmination of a long gradual approach, going back at least to the theory of binary quadratic forms and the demonstration by Jacobi of the functional equations for his θ-functions by means of the Poisson summation formula [**Lang1**, **Rade**]. In recent years, the intimacy between the two subjects was strongly fostered by a constellation of conjectures figured by R. P. Langlands [**Lgld3**, **6**], which, supplemented, refined, and amended by various followers, are generally known under the rubric "Langlands program." These conjectures envision a vast web of "reciprocity laws" (the term comes from quadratic reciprocity through Artin reciprocity) linking "geometric objects" (naively, algebraic varieties defined over $\mathbf{Q}$ or another number field; more sophisticatedly schemes over $\mathbf{Z}$; or now "motives," a rather more elusive notion) with "automorphic representations," the representation-theoretic refinement of the notion of automorphic form, by means of various classes of L-functions. Expositions or developments at several levels of the Langlands program have been published in recent years [**Borl5**, **Gelb**, **ArCl**, **Roga1**, **ClMi**, **Lgld9**].

Here we would like to explain how the study of automorphic forms motivates broadening the purview of representation theory to include not only representations of Lie groups, discussed in §3, but also of Lie group analogs—algebraic groups defined over p-adic (nonarchimedean local) fields.

With some oversimplification, the main problem of the theory of automorphic forms may be said to be the spectral decomposition of $L^2(G/\Gamma)$, where G is a Lie group and $\Gamma \subseteq G$ is a discrete subgroup. That is, we let G act on $L^2(G/\Gamma)$ by left translations, and we want to decompose $L^2(G/\Gamma)$ into (a direct integral of) irreducible representations of G. This formulation is in itself the product of a very substantial conceptual development. It may not appear clearly in traditional presentations **[Lang, Rade, Scho]** of automorphic forms. Originally interest was attached to certain functions on G/Γ with special properties. However, it was gradually realized that many of the key properties of these functions (especially Euler product expansions and criteria for their existence) were naturally expressed in terms of the representation generated by the function and that the function could be retrieved from the representation as a special vector in the space of the representation. This yoga is implicit, for example, in Langlands' (partial) definition of an L-function for every automorphic form **[Lgld3]**. Usually one is interested in the case where Γ is a lattice in G, in fact an arithmetic subgroup (cf. Endnote 4 of §1). We will keep in mind as basic examples the group $\mathrm{SL}_n(\mathbf{Z}) \subseteq \mathrm{SL}_n(\mathbf{R})$, the group of determinant one $n \times n$ matrices with integer entries inside the group of determinant one $n \times n$ real matrices, and $\mathrm{Sp}_{2n}(\mathbf{Z}) \subseteq \mathrm{Sp}_{2n}(\mathbf{R})$, the group of symplectic $2n \times 2n$ matrices with integer entries inside the group of all real $2n \times 2n$ symplectic matrices.

A key feature of arithmetic groups is that they come in families, defined by congruence conditions. Thus, if $\Gamma_1 = \mathrm{SL}_n(\mathbf{Z})$ or $\mathrm{Sp}_{2n}(\mathbf{Z})$ or other arithmetic group, and m is any positive integer, we define Γ_m, the mth *principal congruence subgroup* to be the subgroup of elements γ of Γ_1 which are congruent to 1 (the identity matrix) modulo m, in the sense that the entries of $\gamma - 1$ are all divisible by m:

$$\Gamma_m = \{\gamma \in \Gamma_1 : (\gamma - 1) \equiv 0 \mod m\}. \tag{4.2.1}$$

More generally, a *congruence subgroup* $\Gamma \subseteq \Gamma_1$ is any subgroup which contains Γ_m for some m. (Clearly, any congruence subgroup of Γ_0 will have finite index in Γ_1.) The *congruence subgroup problem* asks when the converse is true: when is any subgroup of finite index in Γ_1 a congruence subgroup? It is often, in fact usually, the case, or almost the case **[Bak, BaMS, Ragh2, PrRa]**. But it, like rigidity (cf. §1.5.2), fails for $\mathrm{SL}_2(\mathbf{R})$. This failure makes $\mathrm{SL}_2(\mathbf{R})$ useful for a variety of problems, including the realization of Galois groups (cf. **[FeFo, Frie, Matz, Thom]**, etc.).

The theory of automorphic forms asks not simply about $L^2(G/\Gamma_0)$ for a fixed arithmetic subgroup Γ_0, but wishes to describe $L^2(G/\Gamma)$ for all congruence subgroups of Γ. Thus in the classical theory of modular forms, one speaks of the *level* of a modular form; a rough translation to the context of our discussion would take "level" to mean the smallest n such that a given function (or the representation it generates) lives on G/Γ_n.

If $\Gamma'' \subset \Gamma' \subseteq \Gamma_1$ are two arithmetic subgroups, then there is an obvious inclusion

$$L^2(G/\Gamma') \hookrightarrow L^2(G/\Gamma''),$$

so we may consider the union or inductive limit

$$\mathscr{L}(G, \Gamma_1) = \bigcup_{\Gamma \subseteq \Gamma_1} L^2(G/\Gamma) \tag{4.2.2}$$

over all arithmetic subgroups $\Gamma \subseteq \Gamma_1$. The theory of automorphic forms aspires to describe $\mathscr{L}(G, \Gamma_1)$, including its level structure, namely where a given representation sits in the hierarchy of $L^2(G/\Gamma)$ inside $\mathscr{L}(G, \Gamma_1)$.

It was Hecke **[Heck]** who first observed (in a context where the issues were considerably more obscure) that $\mathscr{L}(G, \Gamma_1)$ allows substantially more symmetry than just the action of G by left translations. Implicit in the definition of Γ_1 is a group $G_{\mathbf{Q}}$, the rational points of G. We have $\Gamma_1 \subseteq G_{\mathbf{Q}} \subseteq G$, and $G_{\mathbf{Q}}$ is dense in G. Thus, if $\Gamma_1 = \mathrm{SL}_n(\mathbf{Z})$ and $G = \mathrm{SL}_n(\mathbf{R})$, then $G_{\mathbf{Q}} = \mathrm{SL}_n(\mathbf{Q})$; and similarly for the example of the symplectic group. The group $G_{\mathbf{Q}}$ has the property that for any g in $G_{\mathbf{Q}}$, and any congruence subgroups Γ', Γ'' in Γ_1, the intersection $(g\Gamma' g^{-1}) \cap \Gamma''$ is again a congruence subgroup. (The reader should find this easy to verify for $\Gamma_1 = \mathrm{SL}_n(\mathbf{Z})$, $G_{\mathbf{Q}} = \mathrm{SL}_n(\mathbf{Q})$.) We should note that $\mathscr{L}(G, \Gamma_1)$ in fact depends only on $G_{\mathbf{Q}}$, not on a particular arithmetic subgroup Γ_1. Because of this, we will write $\mathscr{L}(G, G_{\mathbf{Q}})$.

Consider $f \in L^2(G/\Gamma)$. For $g \in G_{\mathbf{Q}}$, consider the right translate of f by g:

$$R_g(f)(x) = f(xg), \qquad x \in G.$$

If $\gamma \in \Gamma \cap g\Gamma g^{-1}$, then

$$R_g(f)(x\gamma) = f(x\gamma g) = f(xg(g^{-1}\gamma g)) = f(xg) = R_g(f)(x).$$

Hence $R_g(f)$ belongs to $L^2(G/(\Gamma \cap g\Gamma g^{-1}))$. We have shown the following:

LEMMA 4.2.3. *Right translation by $g \in G_{\mathbf{Q}}$ preserves $\mathscr{L}(G, G_{\mathbf{Q}})$.*

It is obvious that R_g commutes with the action of G on $\mathscr{L}(G, G_{\mathbf{Q}})$ by left translations. Thus $\mathscr{L}(G, G_{\mathbf{Q}})$ is actually a $(G \times G_{\mathbf{Q}})$-module. We may go further. We observe that the action of $G_{\mathbf{Q}}$ is of a special sort: any given f in $\mathscr{L}(G, G_{\mathbf{Q}})$ is stabilized by some congruence subgroup $\Gamma \subseteq \Gamma_1$. This makes reasonable, if it does not directly suggest, the following construction which extends the action of $G_{\mathbf{Q}}$ to a larger group, obtained by a certain process of completion. Define a topology on $G_{\mathbf{Q}}$ by considering the congruence subgroups $\Gamma \subseteq \Gamma_1$ to be open subgroups of $G_{\mathbf{Q}}$. It is easy to check that this makes $G_{\mathbf{Q}}$ into a Hausdorff topological group [**HeRo**, **Loom**, **MoZi2**]. A topological group has a natural sense of uniform structure, hence of Cauchy sequence: a sequence $\{\gamma_j\}_{j=1}^{\infty}$ is Cauchy, provided $\gamma_j\gamma_k^{-1}$ converges to the identity. This is the same as saying $\gamma_j\gamma_k^{-1} \in \Gamma_m$ for any m and all sufficiently large j, k. Denote by $G_{\mathbf{A}_f}$ the completion [**Kell**] of $G_{\mathbf{Q}}$ with respect to this uniform structure. Then, because of the nature of the action of $G_{\mathbf{Q}}$ on $\mathscr{S}(G, G_{\mathbf{Q}})$, as noted above, it is easy to check that the action of $G_{\mathbf{Q}}$ on $\mathscr{S}(G, G_{\mathbf{Q}})$ extends continuously to an action of $G_{\mathbf{A}_f}$. Thus $\mathscr{S}(G, G_{\mathbf{Q}})$ is a $(G \times G_{\mathbf{A}_f})$-module.

To make this statement more concrete, let us examine the structure of $G_{\mathbf{A}_f}$. First, look at the topology of $G_{\mathbf{A}_f}$. By definition, the closure $\overline{\Gamma}$ in $G_{\mathbf{A}_f}$ of a congruence subgroup Γ will be open in $G_{\mathbf{A}_f}$. Since the congruence subgroups of Γ_1 all have finite index in Γ_1, it is easy to see that any sequence $\{\gamma_g\}$ in Γ will have a convergent subsequence in $G_{\mathbf{A}_f}$. In other words, $\overline{\Gamma}$ will be open and compact in $G_{\mathbf{A}_f}$. It is elementary to check that an open subgroup of a topological group is also closed. Thus $G_{\mathbf{A}_f}$ has a system $\{\overline{\Gamma}\}$ of compact, open, and (hence) closed subgroups which form a neighborhood base for the identity in $G_{\mathbf{A}_f}$. In other words, $G_{\mathbf{A}_f}$ is a totally disconnected, locally compact group.

Look also at the algebraic structure of $G_{\mathbf{A}_f}$. For each m, the completion $\overline{\Gamma}_m$ of Γ_m will be a normal, open subgroup of $\overline{\Gamma}_1$, and we will have

$$\overline{\Gamma}_1/\overline{\Gamma}_m \simeq \Gamma_1/\Gamma_m.$$

A review of the definition of $\overline{\Gamma}$ reveals that it may be regarded as an inverse limit [**HeRo**, **KeNa**, **Lang3**] of the Γ_1/Γ_m:

$$\overline{\Gamma} = \varprojlim \Gamma_1/\Gamma_m. \tag{4.2.4}$$

To get a feel for the structure of $\overline{\Gamma}$, consider the case of

$$N = \left\{ \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} : x \in \mathbf{R} \right\}$$

and its subgroup Γ_1, of integral matrices, which is isomorphic to $\mathbf{Z}$. We see then that $\Gamma_1/\Gamma_m \simeq \mathbf{Z}/m\mathbf{Z}$, so that

$$\overline{\Gamma}_1 \simeq \varprojlim \mathbf{Z}/m\mathbf{Z}.$$

The Chinese Remainder Theorem [**Jaco2**, **Lang3**] tells us that if $m = \prod_p p^{j_p}$ is the prime factorization of m, then

$$\mathbf{Z}/m\mathbf{Z} \simeq \prod_p \mathbf{Z}/(p^{j_p}\mathbf{Z}).$$

It follows in the inverse limit that $\overline{\Gamma}_1 \simeq \prod_p \overline{\mathbf{Z}}_p$, where

$$\overline{\mathbf{Z}}_p = \varprojlim \mathbf{Z}/(p^n\mathbf{Z}). \tag{4.2.5}$$

The group $\overline{\mathbf{Z}}_p$, which, in fact, inherits from $\mathbf{Z}$ the structure of ring as well as additive group, is known as the *p-adic integers*. It is an integral domain: if $x \in \mathbf{Z} - p^k\mathbf{Z}$, and $y \in \mathbf{Z} - p^l\mathbf{Z}$, then $xy \notin p^{k+l}\mathbf{Z}$. It has a unique prime ideal $p\overline{\mathbf{Z}}_p$. Its field of quotients is called $\mathbf{Q}_p$, the *p-adic numbers*. It is a non-discrete, locally compact field—for short, a *local field*. The embedding $\mathbf{Z} \to \overline{\mathbf{Z}}_p$ extends to an embedding $\mathbf{Q} \to \mathbf{Q}_p$ with dense image. Thus $\mathbf{Q}_p$ may be regarded as a *completion* of $\mathbf{Q}$: a local field in which $\mathbf{Q}$ embeds densely. A classical result [Weil2] asserts that the fields $\mathbf{Q}_p$, the p-adic numbers for a prime p, and $\mathbf{R}$, the reals, constitute all possible completions of $\mathbf{Q}$. More generally, any local field of characteristic zero is either $\mathbf{R}$, $\mathbf{C}$, or a finite extension of $\mathbf{Q}_p$ for some p [Weil2].

The product

$$\mathbf{Z}_{\mathbf{A}_f} = \prod_p \overline{\mathbf{Z}}_p \tag{4.2.6}$$

over all primes p of the p-adic integers is, like the individual factors $\overline{\mathbf{Z}}_p$, a compact ring. It is not, of course, a domain, because the product of two elements from different factors will be zero. However, it contains $\mathbf{Z}$ (embedded diagonally) as a dense subring, and if we invert elements of $\mathbf{Z}$ we obtain a ring $\mathbf{A}_f$, the ring of *finite adeles*. Since a given $m \in \mathbf{Z}$ is invertible in $\overline{\mathbf{Z}}_p$ if and only if p does not divide m, we can see that $\mathbf{A}_f$ may be described as follows. For every finite set $S = \{p_1, p_2, \dots, p_s\}$ of primes, let $\mathbf{A}_S$ be the ring obtained from $\mathbf{Z}_{\mathbf{A}_f}$ by inverting the primes in S. It is easy to see that

$$\mathbf{A}_S \simeq \left(\prod_{p\in S} \mathbf{Q}_p\right) \times \prod_{p \notin S} \overline{\mathbf{Z}}_p. \tag{4.2.7}$$

One has further that

$$\mathbf{A}_f = \bigcup_S \mathbf{A}_f. \tag{4.2.8}$$

Another description, easily checked to be equivalent to (4.2.8), is that $\mathbf{A}_f$ is the *restricted direct product* of the $\mathbf{Q}_p$ with respect to the $\overline{\mathbf{Z}}_p$: the set of sequences $(x_2, x_3, x_5, \dots)$, where $x_p \in \mathbf{Q}_p$, and, for almost all (in the sense: all but a finite number) p, we have $x_p \in \overline{\mathbf{Z}}_p$. The topology on $\mathbf{A}_f$ is such that each $\mathbf{A}_S$ is an open subring. In particular, the ring $\mathbf{Z}_{\mathbf{A}_f}$ is open in $\mathbf{A}_f$, and $\mathbf{A}_f$ is a locally compact ring.

The ring $\mathbf{A}_f$ of finite adeles can be used to describe the results of our completion construction for a general group $G_{\mathbf{Q}}$, as we anticipated by using the notation G_{A_f} for this completion. Thus for $G = \mathrm{SL}_n(\mathbf{R})$, $G_{\mathbf{Q}} = \mathrm{SL}_n(\mathbf{Q})$, and $\Gamma_1 = \mathrm{SL}_n(\mathbf{Z})$, one has $\overline{\Gamma}_1 = \prod_p \mathrm{SL}_n(\overline{\mathbf{Z}}_p)$ and $G_{A_f} = \mathrm{SL}_n(\mathbf{A}_f)$, which can be described either as the group of matrices of determinant 1 with coefficients in $\mathbf{A}_f$, or as the restricted direct product of the groups $\mathrm{SL}_n(\mathbf{Q}_p)$ with respect to the open compact subgroups $\mathrm{SL}_n(\mathbf{Z}_p)$. With some caveats, this is the general pattern: if $G_{\mathbf{Q}}$ is described as a group of matrices with rational entries satisfying some equations, then $G = G_{\mathbf{R}}$ is the group of matrices satisfying the same equations, but with real numbers as entries; the $G_{\mathbf{A}_f}$ are the matrices satisfying these equations, but with entries in $\mathbf{A}_f$; and, up to finite index, $\overline{\Gamma}_j$ is the product of the $G(\overline{\mathbf{Z}}_p)$.

Let us summarize the consequences of this discussion for the structure of $\mathscr{L}(G, G_{\mathbf{Q}})$. We see we have found that $\mathscr{L}(G, G_{\mathbf{Q}})$ is a module not just for the Lie group $G = G_{\mathbf{R}}$ but for a product group $G_{\mathbf{R}} \times G_{\mathbf{A}_f}$, where the factor $G_{\mathbf{A}_f}$ is itself a (restricted) product of groups $G_{\mathbf{Q}_p}$, for all primes p. The factors $G_{\mathbf{Q}_p}$ of $G_{\mathbf{A}_f}$ look "just like" $G_{\mathbf{R}}$, in the sense that they are a group of matrices satisfying the same equations as the equations defining $G_{\mathbf{Q}_p}$, only the entries of the matrices are in $\mathbf{Q}_p$ rather than $\mathbf{R}$.

The previous paragraph suggests that $G_{\mathbf{R}}$ and the $G_{\mathbf{Q}_p}$ are on an essentially equal footing as far as $\mathscr{L}(G, G_{\mathbf{Q}})$ is concerned. This is so. We give a slight reformulation of the previous paragraph which emphasizes this viewpoint. Set

$$\mathbf{A} = \mathbf{R} \times \mathbf{A}_f. \tag{4.2.9}$$

This is called the ring of *adeles*. Since $\mathbf{R}$ is a completion of $\mathbf{Q}$, just as are the $\mathbf{Q}_p$, we have the diagonal embedding $\mathbf{Q} \to \mathbf{A}$. Slight extension of the discussion of $\mathbf{A}_f$ **[Lang4, Weil2]** shows that

(4.2.10)
(i) $\mathbf{Q} \cap \left(\mathbf{R} \times \prod_p \overline{\mathbf{Z}}_p\right) = \mathbf{Z}$; hence
(ii) $\mathbf{Q}$ is discrete in $\mathbf{A}$, and
(iii) $\mathbf{A}/\mathbf{Q}$ is compact.

We can also form $G_{\mathbf{A}} = G_{\mathbf{R}} \times G_{\mathbf{A}_f}$. Under often satisfied assumptions on G **[Pras, Plat]**, analogs of facts (4.2.10) hold for G also.

(4.2.11)
(i) $G_{\mathbf{Q}} \subseteq G_{\mathbf{A}}$ is a discrete subgroup.
(ii) The quotient space $G_{\mathbf{Q}}/G_{\mathbf{A}}$ has finite volume.
(iii) If $\Gamma \subseteq G_{\mathbf{Q}}$ is an arithmetic group, and $\overline{\Gamma}$ is the completion of Γ in $G_{\mathbf{A}_f}$, then $\overline{\Gamma}$ is compact, and open in $G_{\mathbf{A}_f}$. Further $G_{\mathbf{R}} \times \overline{\Gamma}$ is open in $G_{\mathbf{A}}$, and $G_{\mathbf{Q}} \cap (G_{\mathbf{R}} \times \overline{\Gamma}) = \Gamma$.
(iv) $G_{\mathbf{R}} \cdot G_{\mathbf{Q}}$ is dense in $G_{\mathbf{A}}$.

Property (iv) is known as *strong density* **[Pras, Plat]**. It follows from points (iii) and (iv), that the inclusion $G_{\mathbf{R}} \hookrightarrow G_{\mathbf{A}}$ gives rise to an identification

$$G_{\mathbf{R}}/\Gamma \simeq \overline{\Gamma}\backslash G_{\mathbf{A}}/G_{\mathbf{Q}}. \tag{4.2.12a}$$

Observe that since $G_{\mathbf{R}}$ commutes with $\overline{\Gamma}$, it will act on $\overline{\Gamma}\backslash G_{\mathbf{A}}/G_{\mathbf{Q}}$ on the left, and (4.2.12a) is a $G_{\mathbf{R}}$-equivariant identification. Of course (4.2.12a) leads to an identification of L^2 spaces:

$$L^2(G_{\mathbf{R}}/\Gamma) \simeq L^2(\overline{\Gamma}\backslash G_{\mathbf{A}}/G_{\mathbf{Q}}). \tag{4.2.12b}$$

All the spaces $L^2(\overline{\Gamma}\backslash G_{\mathbf{A}}/G_{\mathbf{Q}})$ may be considered as subspaces of $L^2(G_{\mathbf{A}}/G_{\mathbf{Q}})$. Doing this, we see that the inclusions $L^2(G_{\mathbf{R}}/\Gamma') \subseteq L^2(G_{\mathbf{R}}/\Gamma'')$ when $\Gamma'' \subseteq \Gamma'$ are simply inclusions of subspaces of $L^2(G_{\mathbf{A}}/G_{\mathbf{Q}})$. Thus the space $\mathscr{L}(G, G_{\mathbf{Q}})$ of equation (4.2.2) is also seen as a subspace of $L^2(G_{\mathbf{A}}/G_{\mathbf{Q}})$—precisely the subspace of all vectors which are fixed by some open subgroup of $G_{\mathbf{A}_f}$. Clearly $\mathscr{L}(G, G_{\mathbf{Q}})$ is dense in $L^2(G_{\mathbf{A}}/G_{\mathbf{Q}})$; or from the point of view of (4.2.2), $L^2(G_{\mathbf{A}}/G_{\mathbf{Q}})$ is the result of taking the Hilbert space completion of $\mathscr{L}(G, G_{\mathbf{Q}})$—in some sense it is the result of following construction (4.2.2) to its natural end. For groups not satisfying strong density (property (4.2.11)(iv)), the connection between $L^2(G/\Gamma)$ and $L^2(G_{\mathbf{A}}/G_{\mathbf{Q}})$ is more complicated than (4.2.12b), but the adelic viewpoint is still illuminating.

In summary, we find that, if we are interested in describing the spaces $L^2(G_{\mathbf{R}}/\Gamma)$ for all congruence subgroups Γ of the arithmetic subgroup Γ_1 and the relations between these spaces, then essentially we are interested in $L^2(G_{\mathbf{A}}/G_{\mathbf{Q}})$. This space supports an action not simply of $G_{\mathbf{R}}$ but of the adele group $G_{\mathbf{A}}$, in which each p-adic completion $G_{\mathbf{Q}_p}$ of $G_{\mathbf{Q}}$ participates on an equal footing with $G_{\mathbf{R}}$. The unitary dual of $G_{\mathbf{A}}$ is identifiable to the (restricted) *product* of the unitary duals of the local factors $G_{\mathbf{Q}_p}$, via a tensor product construction **[Flat]**. Thus an irreducible $G_{\mathbf{A}}$-subspace of $L^2(G_{\mathbf{A}}/G_{\mathbf{Q}})$—which is called an *automorphic representation*—carries vastly more information than simply the $G_{\mathbf{R}}$-isomorphism class that it defines: it also determines a point in $\widehat{G}_{\mathbf{Q}_p}$ for all primes p. Langlands **[Lgld3]** has shown how to use the parameters from all the $\widehat{G}_{\mathbf{Q}_p}$ to define (for almost all primes) local Euler factors, to be multiplied together to form an L-function (the Langlands automorphic L-function) to be attached to the automorphic representation. One can then hope that this automorphic L-function is equal to an L-function attached to some geometric object, which equality would constitute a reciprocity law of some sort. A large number of workers have been engaged in this project, and have made notable progress **[GeSh, JaLa, ArCl, Roga, ClMi, GePs]**, including the establishment of some nonobvious cases of the Artin conjecture on holomorphicity of L-functions **[Lgld5, Tunn]**. However, in some sense, the work has only begun.

Remark. At this point, we should perhaps deal with an issue that may have been bothering the reader, namely what if we do indeed wish to be very classical and deal with $L^2(G/\Gamma)$ for fixed Γ. Individual elements of $G_{\mathbf{A}_f}$ do not preserve individual spaces $L^2(G/\Gamma)$, only the system $\mathscr{L}(G, G_{\mathbf{Q}})$ formed from all $L^2(G/\Gamma)$. However, when we make the identification $L^2(G/\Gamma) \simeq L^2(\overline{\Gamma}\backslash G_{\mathbf{A}}/G_{\mathbf{Q}})$, we see that the convolution algebra $C_c(G_{\mathbf{A}_f}//\overline{\Gamma})$, of compactly supported functions on $G_{\mathbf{A}_f}$ which are both left and right invariant under $\overline{\Gamma}$, will preserve $L^2(\overline{\Gamma}\backslash G_{\mathbf{A}}/G_{\mathbf{Q}})$. Interpreted in terms of $L^2(G/\Gamma)$, elements of $C_c(G_{\mathbf{A}_f}//\overline{\Gamma})$ are the original Hecke operators **[Heck, Lang5, Shim1]**. For this reason, $C_c(G_{\mathbf{A}_f}//\overline{\Gamma})$, is sometimes called a *Hecke algebra.* If $\overline{\Gamma} = \prod \overline{\Gamma}_p$ is a product of local factors, then also $C_c(G_{\mathbf{A}_f}//\overline{\Gamma})$ can be decomposed as a (restricted) tensor product

(4.2.13)
$$C_c(G_{\mathbf{A}_f}//\overline{\Gamma}) \simeq \bigotimes_p C(G_{\mathbf{Q}_p}//\overline{\Gamma}_p)$$

of "local" Hecke algebras $C(G_{\mathbf{Q}_p}//\overline{\Gamma}_p)$. These local Hecke algebras (also known as spherical function algebras) are very interesting. They control the representation theory of $G_{\mathbf{Q}_p}$ **[HoMy1, 2, BDKM, BuKu]**. Also for a particular choice of $\overline{\Gamma}_p$ (the Iwahori subgroup **[Iwah2, Roga2]**), they are related to quantum groups, knot theory, etc. **[Cher2, GoHJ]** (cf. the articles by Jones and Witten in this volume).

The adelic formulation of the theory of automorphic forms raises the question of understanding representations of $G_{\mathbf{Q}_p}$ as well as representations of $G_{\mathbf{R}}$. The issue of representations of p-adic groups has been studied since the early 1960s. Considerable progress has been made, but substantial mysteries remain: the situation is not nearly so complete as in the real case. The construction of principal series by means of induction from parabolic subgroups works as for F, and there is a parallel of "Langlands classification" (Theorem 3.6.4.5) **[Silb, BeZe]**. However, the essence of Theorem 3.6.4.5 is that it describes general admissible representations in terms of discrete series, and the discrete series for p-adic groups remain obscure. It is roughly true that, as for real groups, discrete series are attached to characters of compact Cartan subgroups. However, in the p-adic case there are typically many compact Cartan subgroups instead of at most one, as for real groups, and understanding precisely how they interact has been difficult. The group GL_2, and closely related groups, have been under fairly good control since the 1960s, **[Sall, Shal, Silb, GGPS]**, although the case of residual characteristic 2 gave considerable trouble **[Kutz, Tunn1]**. Some other cases, where the geometry of the compact tori is relatively simple or well behaved, have been treated more or less completely **[CoHo, HoMy1, 2, Moy1, 2, KoZi]**. In recent significant progress, constructions of all discrete series for $\mathrm{GL}_n(\mathbf{Q}_p)$ have been devised **[BuKu, Corw1, 2, Henn2]**. In **[BuKu]**, as in **[HoMy1, 2]**, the main point is to find many copies of the Iwahori Hecke algebras (**[Iwah2,]**

Roga2) in $C_c(G_{\mathbf{Q}_p})$. The representations of Iwahori Hecke algebras have been described in [**KaLu2**].

In fact, Langlands [**Lgld2, 3**] proposed a parametrization of representations of $G_{\mathbf{Q}_p}$ (or $G_{\mathbf{R}}$) in terms of the Galois theory of $\mathbf{Q}_p$. His ideas have been useful even for real (i.e., Lie) groups—they provide the rationale for the "Langlands classification" (Theorem 3.6.4.5), and the duality phenomenon that they suggest provides a deep organizational principle for representations of different real forms of the same complex group [**Voga8, AdVo**]. His proposal, which has gone through several refinements [**Arth2, 3, Borl5, Lusz5**], and may well go through more, can be thought of as a nonabelian generalization of the local reciprocity law of local classfield theory [**Lang4, Weil2, Lgld2**].

To explain the gist of this idea, which applies more or less uniformly to $G_{\mathbf{R}}$ and $G_{\mathbf{Q}_p}$, let us adopt a more neutral notation, and denote by F either $\mathbf{R}$ or $\mathbf{Q}_p$. (In fact, the discussion below will apply equally well for F a local field of characteristic zero.) By G_F we will mean (the F-rational points of) an (affine) algebraic group defined over F (cf. §1.5.2, especially Endnote 4, also [**Borel, Jant, Spri2**], etc.); the reader may think of $\mathrm{SL}_n(F)$, $\mathrm{Sp}_{2n}(F)$, etc. Also, let us remark that, just as for real reductive groups (cf. §3.6.5 and [**Knap2, Wall2**]) there is for p-adic reductive groups a notion of *admissible representation* [**JaLa, Cart2**], with analogous properties: the admissible representations contain the irreducible unitary representations [**Bern 4, Cart 2**], and constitute a sort of analytic continuation of them. Denote the set of irreducible admissible representations of G_F by $\mathrm{Adm}(G_F)$.

The core of Langlands' idea is that irreducible representations of G_F should be parametrized "naturally" by representations of $\mathrm{Gal}(F)$, the (absolute) Galois group of F (which with its standard (inverse limit) topology is thought of as a compact, totally disconnected group [**Lang3**, p. 351]). To be useful, this idea needs considerable clarification and qualification.

The most important qualification regards the nature of the representations of $\mathrm{Gal}(F)$ used to parametrize $\mathrm{Adm}(G_F)$. Since the sets $\mathrm{Adm}(G_F)$ vary considerably for varying G_F, it is not reasonable to parametrize all $\mathrm{Adm}(G_F)$ by the same representations of $\mathrm{Gal}(F)$—one needs a way of segregating the representations of $\mathrm{Gal}(F)$ which should be associated to $\mathrm{Adm}(G_F)$. Langlands' proposal for doing this is to construct a complex Lie group ${}^LG^0$, the *L-group* (more correctly, its identity component), whose structure reflects that of G_F. The construction of ${}^LG^0$ exploits a kind of duality in the family of semisimple groups [**Lgld3, 4, Borl5**]. The existence of this duality emerges from the abstract specification of semisimple groups in terms of Weyl groups and root systems, refining Theorem 2.12.2 [**Spri2, Crtr**]. In this duality, symplectic groups are matched with odd-dimensional orthogonal groups, while all other members of the Killing-Cartan classification (cf. §2.10) are matched with themselves on the Lie algebra level; however, centers also change with duality: SL_n is dual to PGL_n. However, GL_n is dual to itself. Langlands suggests that, rather than look at representations in the usual sense of $\mathrm{Gal}(F)$, which are homomorphisms to $\mathrm{GL}_m(\mathbf{C})$ for some m, we should consider homomorphisms to ${}^LG^0$ to parametrize $\mathrm{Adm}(G_F)$.

The second key qualification of the basic idea is that the Galois group is too confined to correctly reflect $\mathrm{Adm}(G_F)$, because as one sees from parabolic induction, there are continuous families in $\mathrm{Adm}(G_F)$, whereas $\mathrm{Gal}(F)$, being compact, will have only a discrete set of representations. Here one takes a hint from abelian class field theory **[Weil2, Lang4]**, and replaces $\mathrm{Gal}(F)$ by $W(F)$, the *Weil group* of F **[Tate, ArTa]**. When F is an extension of $\mathbf{Q}_p$, this is a group which fits in a diagram

$$\begin{array}{ccc} W(F) & \longrightarrow & F^{\times} \\ \downarrow & & \downarrow \\ \mathrm{Gal}(F) & \longrightarrow & \mathrm{Gal}(F)^{\mathrm{ab}} \end{array} \qquad (4.2.14a)$$

where $\mathrm{Gal}(F)^{\mathrm{ab}}$ is the maximal abelian quotient of $\mathrm{Gal}(F)$. The vertical maps are injections, and the right-hand one is the reciprocity map of abelian classfield theory (cf. **[Weil2, Lang4]**, etc.). Thus $W(F)$ is a loosened-up version of $\mathrm{Gal}(F)$: it is a dense subgroup, equipped with a stronger topology so that it is again locally compact. For $\mathbf{R}$ and $\mathbf{C}$, one has, by definition

$$\begin{gathered} W(\mathbf{C}) = \mathbf{C}^{\times}, \\ 1 \to \mathbf{C}^{\times} \to W(\mathbf{R}) \to \mathbf{Z}/2\mathbf{Z} \to 1, \end{gathered} \qquad (4.2.14b)$$

where the nontrivial element of $W(\mathbf{R})/\mathbf{C}^{\times}$ acts on $\mathbf{C}^{\times}$ by complex conjugation, and has square equal to -1 in $\mathbf{C}^{\times}$. For $\mathbf{R}$ and $\mathbf{C}$, it is especially important to replace the Galois group, which is only of order 2 for $\mathbf{R}$ and trivial for $\mathbf{C}$, by the much larger Weil group.

The main part of Langlands' proposal, then, is to parametrize $\mathrm{Adm}(G_F)$ by homomorphisms from $W(F)$ to ${}^LG^0$. Further modifications are still necessary: ${}^LG^0$ must be extended by $W(F)$, and the homomorphisms restricted in various ways, in order to better reflect the structure of G_F, as opposed to $G_{\widetilde{F}}$, where $\widetilde{F}$ is the algebraic closure of F. Thus we want to be able to distinguish between $\mathrm{GL}_m(\mathbf{R})$ and the unitary groups $\mathrm{U}_{p,q}$, $p+q=m$. Also some extra data is needed, provided by using the Weil-Deligne group **[Tate]** instead of the Weil group, or adding a nilpotent element in the Lie algebra of ${}^LG^0$ to the parametrizing data, or other means **[Borl5]**. The end of this process is a set $\Phi(G_F)$ of "admissible homomorphisms" from $W(F)$ to ${}^LG^0$, and these should parametrize $\mathrm{Adm}(G_F)$ up to finite ambiguity. That is, there should be a map

$$\begin{array}{c} \mathrm{Adm}(G_F) \\ \big\downarrow \\ \Phi(G_F) \end{array} \tag{4.2.15}$$

with finite fibers. The fibers are called *L-packets*, and two representations in the same fiber are called *L-indistinguishable*, to suggest there is no way to separate them using the theory of L-functions. This map should satisfy various desiderata [**Borl5**].

Although the definition of the map (4.2.15) is quite involved for general G, in special cases it has immediate intuitive impact. For example, for $G_F = \mathrm{GL}_n(\mathbf{Q}_p)$, the map (4.2.15) amounts to a canonical bijection between the sets

(i) irreducible n-dimensional complex representations of $W(\mathbf{Q}_p)$,
(ii) irreducible supercuspidal representations of $\mathrm{GL}_n(\mathbf{Q}_p)$.

Supercuspidal representations are a remarkable phenomenon of p-adic representation theory: they are discrete series representations whose coefficients are not merely square-integrable, but have compact support (!). For a group like GL_n, with a noncompact center, terms like "square-integrable" or "completely supported" must be understood "modulo the center" [**Cart2**].

A parametrization of type (4.2.15) is known to exist for GL_l if l is prime, over all p-adic fields [**Kutz, KuMo, Tunn1, Henn1**], and for $\mathrm{GL}_n(F)$ if n is less than p, when F is an extension of $\mathbf{Q}_p$ [**Moy**]. It is also known for all real groups [**Lgld4, Borl5**]. This is the true "Langlands classification." Theorem 3.6.4.5 is a bowdlerized version, expressed solely in terms of the structure of $G_{\mathbf{R}}$, with reference to $\Phi(G_{\mathbf{R}})$ expunged.

L-indistinguishability is connected with a concrete phenomenon in the theory of automorphic forms, a phenomenon which leads to tremendous complications: the failure of "strong multiplicity one." We have noted that a representation Π of the adele group $G_{\mathbf{A}}$ is essentially constructed as a tensor product of representations Π_p of the local factors $G_{\mathbf{Q}_p}$ of $G_{\mathbf{A}}$. In particular, representations Π and Π' of $G_{\mathbf{A}}$ will be equivalent if and only if Π_p is equivalent to Π'_p for all p (including the case of $G_{\mathbf{R}}$ as "the infinite prime" [**Weil2, Lang4**]). Let us call Π and Π' *nearly equivalent* if Π_p and Π'_p are equivalent for all but a finite number of p. Strong multiplicity one for G says that given two automorphic representations Π and Π' of $G_{\mathbf{A}}$, i.e., constituents of $L^2(G_{\mathbf{A}}/G_{\mathbf{Q}})$, then if Π and Π' are nearly equivalent, they *are* equal. In other words, if we are given representations Π_p of the local factors $G_{\mathbf{Q}_p}$ of $G_{\mathbf{A}}$, for all but a finite number of p, and told to make an automorphic representation out of them, then there is at most one way to choose the remaining Π_p in order to do this and at most one way to put the resulting representation of $G_{\mathbf{A}}$ in $L^2(G_{\mathbf{A}}/G_{\mathbf{Q}})$. Strong multiplicity one is known to hold for GL_n, and inside large classes of automorphic forms for other groups **[Piat]**. In fact, for GL_n, knowing an automorphic representation at a sufficiently large *finite* number of places is enough to determine it completely **[More]**. However, for many groups, strong multiplicity one fails. Piatetski-Shapiro has been especially vigorous in providing examples of this failure, and of other peculiar phenomena of automorphic forms **[CoPS1, 2, HoPS1, 2]**.

Thus, for a group (such as the symplectic groups) for which strong multiplicity one fails, we can find two distinct but nearly equivalent automorphic representations Π and Π'. Let Π_p and Π'_p be local components of Π and Π'. Then it should be the case that Π_p and Π'_p are L-indistinguishable, i.e., are in the same fiber of the parametrization map (4.2.14). This is the concrete meaning of L-indistinguishability. Langlands and his school have spent much effort in recent years grappling with problems arising from L-indistinguishability **[LaLa, Lgld8, LaSh, Shel]**. Current interest is focussed on Hasse-Weil zeta-functions of Shimura varieties (cf. **[Lgld6, ClMi, Rogal, Miln]**, etc.). A major tool in this endeavor is the trace formula **[Selb2, Arth4, 5, Labe, Rogal]**.

Even in its conjectural state, the parametrization (4.2.15) is of considerable interest, because of the astounding amount of structure it suggests in the representation theory of groups defined over a given local field. It points toward the existence of an intricate interlocking system of correspondences between representations of different groups. Let G_F and H_F be two groups over the local field F. Suppose there is a homomorphism

$$\alpha : {}^L G \to {}^L H$$

between their respective L-groups. If $\varphi : W(F) \to {}^L G$ is a homomorphism from the Weil group to ${}^L G$, then $\alpha \circ \varphi$ will be a homomorphism from $W(F)$ to H. Thus, modulo technicalities, we would expect composition with α to define a mapping

$$\alpha_* : \Phi(G_F) \to \Phi(H_F).$$

The obvious question is whether the mapping α_* somehow lifts to define a commutative square

$$\begin{array}{ccc} \mathrm{Adm}(G_F) & \xrightarrow{\tilde{\alpha}_*} & \mathrm{Adm}(H_F) \\ \downarrow & & \downarrow \\ \Phi(G_F) & \xrightarrow{\alpha_*} & \Phi(H_F) \end{array} \tag{4.2.16}$$

where the vertical maps are as in (4.2.15). Langlands' *Principle of Functoriality* posits the existence of such squares **[Borl5]**. In particular, it posits the existence of maps $\tilde{\alpha}_*$ between irreducible admissible representations of G_F and H_F. We may look for such maps $\tilde{\alpha}_*$ whether or not we know that the vertical maps exists.

Many maps of the form $\tilde{\alpha}_*$ are known to exist. The maps of parabolic or cohomological induction (cf. §§3.6.1, 3.6.4) can be regarded as such maps (corresponding to maps α given by embeddings of Levi components of parabolic subgroups). Similar remarks apply to the correspondences defined by homomorphisms of Hecke algebras (p-adic "Harish-Chandra homomorphisms") constructed in **[HoMy1, 2, BuKu, Wald3]**. Viewing such correspondences as being of the type (4.2.16) in no way simplifies establishing that they exist, but it makes them seem plausible and suggests they are part of a larger pattern.

Other examples are the "base change" mappings between a group over a given field F and the same group but with coefficients in an extension field F' of F. The most frequently studied case is when F' is a cyclic extension of prime degree over F **[ArCl, Lgld5, GeLa]**. This correspondence has been studied globally (i.e., for automorphic forms) as well as locally (for admissible representations) and has resulted, among other things, in the establishment of some new cases of the Artin conjecture **[Lgld5, Tunn2]**. Establishment and use of correspondences suggested by the Principle of Functoriality is becoming standard operating procedure in the theory of automorphic forms **[Lgld5, BlRa, Roga1]**. Here also the trace formula is heavily used.

The theory of θ-series, which is the automorphic aspect of the oscillator representation **[Howe5]** also produces, in a god-given way, correspondences between automorphic forms, made up of local correspondences between admissible representations of certain pairs of groups (G, G') (precisely "reductive dual pairs"—mutually centralizing subgroups of the symplectic group Sp_{2n}) (cf. **[Howe5, Gelb2, MoVW, Moeg, Prze, Mand, Wald, Shim2, Shin2, Niwa]**, etc.). In some cases, these "θ-correspondences" can be shown to be of the form predicted by the Principle of Functoriality. However, for groups over $\mathbf{R}$ (i.e., Lie groups), where the Langlands classification (4.2.15) is known, and which therefore currently offer the strongest test, some θ-correspondences are incompatible with the L-packet parametrization. That is, some θ-correspondences do not preserve L-packets **[Adam1]**.

On the other hand, Arthur **[Arth2, 3]**, influenced by earlier examples of θ-series which violated the "generalized Ramanujan conjecture" **[Sata, HoPS, Kuro]**, and by problems stemming from L-indistinguishability, was led to propose that certain parts of $\mathrm{Adm}(G_F)$ should be coagulated into lumps larger than L-packets. These larger lumps are called ψ-packets. It seems possible that θ-correspondences will be compatible with ψ-packets **[Adam1]**. Further, it appears that Arthur's proposals are directly relevant to the determination of the unitary dual for semisimple Lie groups **[Arth2, 3, BaVo]** (and presumably p-adic groups also). This complex interplay of rich examples, difficult technical issues, and bold ideas makes the theory of automorphic forms an exciting research area, one whose challenges will occupy generations to come.

4.3. *Physics, geometry, and differential equations.* Physics, geometry, and differential equations are intimately related to each other, and the interaction of Lie theory with all of them has been extensive. This is hardly surprising, since all three of geometry, differential equations, and Lie theory are the children of physics, and group theory is now understood to provide a large part of the underpinnings of both geometry and analysis. On a philosophical level, one may also observe that a great deal of physics is concerned with conservation laws and invariance principles, and group theory is a natural language for expressing such ideas. The application of Lie theory to physics is the subject of a long series of large conferences [**ITGT1–17**, **Loeb**] and numerous texts [**BiLo1**, **2**, **BeTu**, **Corn**, **Hame**, **Jone**, **Lich**, **Wolb**], etc. We have given perhaps the most basic example in quantum mechanics, the harmonic oscillator, in §3.1. The energy states of the hydrogen atom, i.e., the quantum Coulomb problem, provide another elegant and subtle example with "accidental degeneracies" accounted for by extra symmetries [**Abar**, **Engl**, **Fron**, **Shan**]. Shortly after Einstein introduced special relativity [**Eins**], Minkowski [**Mink**] explained that the difference between classical (Newtonian or Gallilean) physics and relativity could be explained as a change in the symmetry group of space time, from the isometry group of the degenerate form $x_1^2 + x_2^2 + x_3^2$ in four variables to that of the nondegenerate form $x_1^2 + x_2^2 + x_3^2 - x_4^2$. This is analogous to the passage from Euclidean to non-Euclidean (Lobachevskian) geometry. Several early cosmological models of the expanding universe are based on homogeneous spaces for various semisimple Lie groups [**Eins2**, **Sitt**]. A speculative alternative to "big bang" models of the universe has been proposed by I. Segal [**Sega4**], based on $S^3 \times \mathbf{R}$ as a homogeneous space for $\mathrm{SU}(2, 2)^{\sim}$, where $^{\sim}$ indicates the universal cover. R. Penrose's formulation of general relativity in terms of "twistors" [**PeWa**, **Penr**, **Well**] is based on Lie-theoretic constructions. Modern bookkeeping schemes for elementary particles, beginning with isotopic spin, and continuing with the 8-fold way [**Gü Ra**, **Loeb**, **SaWe**], and beyond, are based on the combinatorics of finite-dimensional representations of Lie groups (SU_2, SU_3, etc.). The Yang-Mills equations, thought to be the governing equations of "quantum chromodynamics," the interactions of quarks, are variational equations on the space of "gauge fields," connections of a principal fiber bundle, whose fiber is an appropriate Lie group [**AtBo**, **FrUh**, **Taub**]. Recently the speculative "string theory" and its cousin conformal field theory [**BePZ**, **Gawe**, **MoSe**, **Segl**, **Witt**], have contributed to the explosive growth in the study of infinite-dimensional Kac-Moody Lie algebras, especially the "affine" Lie algebras (cf. [**Dola**, **FrKa**, **FrLM**, **GrSW**, **Kac1**], etc.). Currently the study of "quantum Lie algebras" or "quantum groups," which grew out of certain formal identities arising in exactly solvable statistical mechanical models [**Andr**, **AnBF**, **Baxt**, **JiMi1**, **2**], is proceeding furiously, and is leading to new insights even in matters of longstanding interest in finite-dimensional representation theory of semisimple Lie groups [**Kash**, **Lusz4**, **Cher1**, **2**, **Murp**].

Some of these developments have led to dramatic new advances in geometry, for example, Donaldson's analysis of 4-manifolds as boundaries of solutions of families of solutions to the Yang-Mills equations [**FrUh**]. Donaldson constructed new invariants of 4-manifolds; there have also been many recent constructions of invariants for 3-manifolds, including [**Cran1**, **2**, **TuVi**], which construct 3-manifold invariants in terms of the quantum $6-j$ coefficients, q-analogs of numbers which arise in the explicit description of tensor products of representations of SL_2 [**BiLo1**, **2**].

The more traditional applications of Lie theory to geometry are also extensive, and many are fundamental. One of the first that can be considered explicitly Lie-theoretic (although it precedes the period of coverage of this essay) is the Erlanger Programm of Felix Klein [**Klei2**], in which the relation between Euclidean geometry and the various alternatives (the hyperbolic geometry of Lobachevsky and Bolyai, the elliptic geometry of Riemann, the projective geometry, etc.) which had arisen in the nineteenth century, are systematically related to one another in terms of their associated symmetry groups; and in which, furthermore, the word "geometry" is proposed to mean the understanding of the invariants of a group G acting on several copies of a homogeneous space. (Thus for the Euclidean group of isometries of the plane, the only invariant of two points is the distance between them; the invariants of three points are side-angle-side, or angle-side-angle, etc.) The variety of possible geometries was the question which led to Killing's [**Kill**] (see also [**Cole**]) classification of the simple Lie algebras, and which received a more or less definitive formulation in E. Cartan's definition and classification of symmetric spaces [**Crtn4**], (see also [**Helg2**, **Loos**]). Similarly, invariant theory, which was more or less explicitly Lie-theoretic in nature even before Lie theory existed, and which led to tremendous computational efforts in the nineteenth century [**Sylv1**, **2**], after being cut off at the root by Hilbert [**Hilb1**, **2**], found new life by being grafted onto the representation theory of semisimple groups [**Litt**, **Weyl2**], and made a basis for the theory of moduli of algebraic varieties [**FoMu**].

These evocations and lists could go on indefinitely. Let us mention just a very few particular examples. Efforts to understand the topology, especially the cohomology ring, of Lie groups [**Crtn5**, **6**], inspired deRham's Theorem [**Rham**], and led to the notions of H-space [**Hopf3**, **Brow**, **Thms**] and Hopf algebra [**Hopf2**, **HoSa**, **Swee**]. This latter notion has been crucial in the recent formulation of the idea of quantum group [**Drin**, **Jimb**]. In Riemannian geometry, a major theme has been the study of spaces of positive (sectional) curvature. Most of the known examples are homogeneous spaces or perturbations of them [**Barg2**, **Wall6**]. In even dimensions, there can be only a finite number of homotopy types of positively-pinched manifolds [**ChEb**], and there are only a finite number of homogeneous examples in each even dimension. However, Wallach [**Wall6**, **AlWa**] classified all possible homogeneous, positively curved manifolds, and found an infinite family of seven-dimensional examples. Another elegant example in a similar spirit is an exotic 7-sphere with nonnegative curvature, constructed as a quotient of Sp_2 by an action of $Sp_1 \simeq SU_2$ built from right and left translations [**GrMe**]. Finally, we mention the question [**KacM**] whether the spectrum of the Laplace-Beltrami operator on functions on a compact Riemannian manifold determines the manifold up to isometry. The answer is "no," a stronger and stronger no as more and more counterexamples have been constructed [**Miln**, **Iked**, **Vign**, **Suna**, **Bera**, **BrGo**]. All these examples are homogeneous spaces, or almost homogeneous spaces, and rely on the structure imposed by group theory to get control of the spectrum of the Laplacian.

Differential equations, especially nonlinear differential equations, were the context for some of Lie's original investigations [**Hawk**, **LiEn**], and certain infinite-dimensional Lie algebras were studied by Cartan [**Crtn5**, **GuSt2**] in connection with variational problems. (Recently, Cartan's methods were used by Olver [**Olve2**] to establish an interesting result in the invariant theory of binary forms and Cartan's algebras, which have finite-dimensional quotients modulo p, figure in the classification of simple Lie algebras over fields of positive characteristic [**StWi**].) For much of the twentieth century, however, such investigations were pursued much less vigorously than ones sparked by the internal development of the subject, but recently, Lie groups as symmetry groups of solutions to differential equations have received renewed attention [**Ibra**, **Olve1**, **Ovsi**].

A relatively recent area, control theory borrows techniques from geometry and differential equations both, and in particular uses Lie theory [**Haze1**, **2**, **Broc1**, **2**, **Crou**, **HaMa**, **Herm1**, **2**]. Being concerned primarily with the future, control theory has stimulated the theory of Lie semigroups [**HiHL**].

Linear partial differential equations have been attacked by methods of harmonic analysis, whose essentially group-theoretic basis was recognized in this century [**PeWe**, **Weil3**, **Weyl1**]. The development of pseudo-differential and Fourier integral operators [**Hörm**] has led to "phase-space analysis" in which the distinction between multiplication and differentiation is blurred. This phase space analysis has a direct interpretation in terms of the Heisenberg group, and many of the basic estimates [**Beal**, **Hörm**, **Unte**] of the theory can be carried out efficiently using group theory, of the Heisenberg group itself, and of the oscillator representation [**Foll**, **Howe3**]. The recent theory of wavelets [**Daub**, **Gros**, **Meye**] to some extent breaks with Lie theory, but still draws inspiration from symmetry principles, especially scaling. Nilpotent Lie groups, especially the Heisenberg group, have been used to establish facts about particular systems of equations, especially hypoellipticity [**Foll2**, **FoSt**, **RoSt**]. The Lewy counterexample [**Lewy**] to local solvability of systems of differential equations has been interpreted as a system of left-invariant operators defining a positive complex polarization (cf. §3.3) on the Heisenberg group **[GrSt]**. Conversely, there has been substantial study of the analytic properties of elements of $\mathscr{U}(g)$ acting on $C^{\infty}(G)$, expecially for G nilpotent (cf. **[Corw3**, **CoHr**, **HeNo**, **Rock]**, etc.).

Further, many of the classical equations of physics have many symmetries, and are otherwise tightly connected to group theory (cf. **[Howe6**, **ITGT**, **Hame**, **Jone]**, etc.). Dirac is said to have been guided to his equation for the dynamics of the electron by a desire to have it be Lorentz invariant. The Laplace operator

$$\Delta = \sum_{j=1}^{n} \frac{\partial^2}{\partial x_j^2}$$

is invariant under the full group of isometries (translations and rotations) of $\mathbf{R}^n$, and is characterized as the generator of the algebra of all differential operators with full Euclidean invariance. In particular, the space of harmonic functions ($\Delta f = 0$) or indeed any eigenspace of Δ, carries a representation of the Euclidean group. The matrix coefficients of these representations are Bessel functions **[Mill1**, **Vile]**. This example may be generalized tremendously. Helgason **[Helg1]** proposes a general program of studying a system consisting of a group G acting on a space X, together with some differential operators $\{D_i\}$ which commute with G and with each other: the goal is to analyze the representations of G defined by the joint eigenspaces of the D_i. Such a study becomes a two way street: one can learn about the eigenspaces if one has control over the representations, or vice versa. Many of the geometric realization theorems, for discrete series and other interesting representations (cf. **[JaVe**, **Mant**, **RaSW**, **PaRo**, **Tora**, **Schm1–3]**, etc.), could be considered examples in an extended version of this program. The most direct examples arise when G is the full group of isometries of a symmetric space X: in that case the full algebra of G-invariant differential operators on X is commutative, and the eigenspaces yield representations infinitesimally equivalent to the spherical principal series (cf. §3.6.1) **[Helg1**, **GaVa]**. In the past decade, there has been a large amount of work generalizing this to the case of semisimple symmetric spaces (cf. **[Ban**, **Bien**, **FlJe1**, **2**, **OlOr**, **Oshi1–3**, **Schl]**, etc.). Matrix coefficients associated to eigenspace representations will yield many classical families of special functions **[GaVa**, **Helg1**, **Mill1**, **Vile]**.

The subject of "dynamical systems" is the modern home for the qualitative theory of differential equations. It combines geometrical, analytical, and dynamical considerations to study a variety of problems, many of which were originally motivated by physics (though in some cases investigations have strayed rather far from their origins). A (discrete time) *dynamical system* in the modern sense is simply a pair (X, T), where X is a set ("space," or "state-space," or "phase space") and $T : X \to X$ is a mapping. We will assume T is invertible, i.e., 1-to-1 and onto, though this is not necessary for all purposes. Usually X will be equipped with structure, e.g., as a measure space, a topological space, a smooth manifold, a metric space, and T preserves that structure. Thus one speaks of measurable dynamics, topological dynamics, smooth dynamics, rigid motions, or other specific kinds of dynamical system, in order to specify context. In all cases, the focus is on motion, which is described by T. Thus one is implicitly studying the cyclic group with generator T. If one wishes to think of continuous motion, one considers a flow or one-parameter group (cf. §1.2.1) of transformations T_r, $r \in \mathbf{R}$, rather than a single T. There is a standard construction (suspension, or "flow under a function," cf. **[Pete]**) to convert a discrete time dynamical system to a flow.

Even at this early stage, before any special structures have been imposed, it is clear that group theory provides a wealth of examples of dynamical systems: any triple (G, H, g), where G is a group, $H \subseteq G$ is a subgroup, and g is an element of G, defines a dynamical system $(g, G/H)$, with g acting on G/H by left translation. If G is a Lie group, we can make flows by considering one-parameter subgroups rather than single elements. We will see below that some of these examples are extremely interesting.

Many questions asked under the rubric of dynamical systems concern the long-term behavior of these systems. *Ergodic theory* studies the *average* long-term behavior of systems (X, T), where X is equipped with a probability measure μ, which T preserves. Ergodic theory was motivated originally by a question in statistical mechanics. To explain this, we recall some basic notions and facts of the subject. Suppose we can write $X = X_1 \cup X_2$, a disjoint union, with both X_1 and X_2 of positive measure, and both invariant under T. Then writing $T_j = T_{|X_j}$, we may consider that (T, X) is composed of two separate, at least in the measure-theoretic sense, subdynamical systems (X_j, T_j), $j = 1, 2$. Interest obviously focuses on systems which are not decomposable in this fashion: these indecomposable systems are called *ergodic*, and a main problem of ergodic theory has been: given a system (X, T) decide whether it is ergodic.

A basic result which brings out the significance of ergodicity is Birkhoff's Ergodic Theorem **[Halm, Pete]**. Let (X, T) be a measurable dynamical system, and denote by μ the given T-invariant measure on X. Given $p \in X$, and a (measurable) function f on X, define the *time averages* along the orbit of p by

$$A_n(f)(p) = \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(p)), \qquad n \geq 1. \tag{4.3.1}$$

THEOREM 4.3.2. *For $f \in L^1(x, \mu)$, there is a function $A_\infty(f)$ in $L^1(X, \mu)$, invariant by T, such that*

$$\lim_{n\to\infty} A_n(f)(p) = A_\infty(f)(p)$$

for almost all p *(with respect to* μ*). In particular, if* T *acts ergodically on* X*, then*

$$A_n(f)(p) \to \int_X f(x)\,d\mu(x) \tag{4.3.3}$$

for almost all p *in* X.

An easy argument convinces one that T is ergodic if and only if the only T-invariant functions in $L^1(X, \mu)$ are the constant functions. Thus the second statement of Theorem 4.3.2 is an easy corollary of the first. It is the source of the interest of ergodicity in statistical mechanics. The time averages $A_n(f)(p)$ of various dynamical quantities f appear in statistical mechanical calculations; but the calculation of $A_n(f)(p)$ in terms of its definition, involving as it does detailed step-by-step knowledge of the dynamics over long periods of time, is formidably difficult. If $A_n(f)(p)$ can be estimated by the single, predeterminable number $\int_X f(x)\,d\mu(x)$, known as the *phase average*, the life of the statistical mechanics theorist is immeasurably simplified. For further discussion of these matters, see **[Mack1]**.

From the Kronecker line **[ArAv**, p. 132; **Casl]** (a dense one-parameter subgroup of the n-dimensional torus $\mathbf{T}^n$), through the extensive study, especially by Hedlund **[Hedl1, 2]**, and Hopf **[Hopf2]**, in the 1920s and 1930s, of the geodesic flow on Riemann surfaces (endowed with the metric of constant curvature) to the current studies by Dani-Margulis **[Dani1, 2, Marg, DaMa1, 2]** and, more or less definitively, by Ratner **[Ratn4]**, of horocycle flows on G/Γ, Γ a lattice in semisimple G, Lie theory has provided important examples of ergodic actions. These examples frequently bear on number theory, especially questions of Diophantine approximation. The relation of the Kronecker line to Diophantine approximation is classical. We will explain shortly a result of Margulis (generalized dramatically by Ratner **[Ratn4]**) which implies a number-theoretic result which had been an open conjecture for sixty years.

In **[FoGe]**, Fomin and Gelfand showed how to use representation theory to establish ergodicity of dynamical systems constructed from homogeneous spaces, or more generally, obtained from selecting one transformation from a larger group of symmetries. This is based on the fact that ergodicity and related properties can be detected using spectral theory. Precisely, consider a measurable dynamical system (X, T), with invariant measure μ. Let $L^2(X, \mu)$ be the L^2-space of μ, and let $L^2(X, \mu)^o$ be the subspace orthogonal to the constant functions. We can use T to define an endomorphism U_T of $L^2(X, \mu)$, in the usual way:

$$U_T(f)(x) = f(T^{-1}(x)), \qquad f \in L^2(X, \mu),\ x \in X. \tag{4.3.4}$$

Ergodicity of T can easily be formulated in terms U_T **[ArAv, Halm]**: for T to be ergodic, U_T should have only the constant functions as fixed vectors, or should have no fixed vectors in $L^2(X, \mu)^o$, or should not have 1 in its point spectrum on $L^2(X, \mu)^o$. Since fixed vectors can be detected by means of matrix coefficients (v is a fixed vector for U_T if and only if the matrix coefficients (cf. §A.1.11)

$$\varphi_{v,v}(n) = \int_X U_T^n(v)(x)\overline{v(x)}\, dx, \qquad n \in \mathbf{Z},$$

are constant as a function of n), we can also formulate the condition of ergodicity in terms of matrix coefficients.

The matrix coefficient formulation of ergodicity suggests several variant conditions stronger than ergodicity. One of the most useful of these is *strong mixing* [**ArAv**, **Halm**, **Pete**] which is the requirement that the matrix coefficients $\varphi_{v,v}(n)$ should decay to zero as $n \to \infty$ if $v \in L^2(X, \mu)^o$. Strong mixing is a technically pleasant property because it makes sense for any group G and any unitary representation ρ of G on a Hilbert space $\mathscr{H}$: for u, v in $\mathscr{H}$, we should require that the matrix coefficient $(\rho(g)u, v)$ decay to zero as g goes to ∞ in G. In other words, the set of g for which $|(\rho(g)u, v)| \geq \varepsilon$ should be compact. A further useful property of strong mixing is that it is clearly inherited by closed subgroups. By interpreting the geodesic flow on a Riemann surface in terms of $SL_2(\mathbf{R})$, and then essentially showing that any nontrivial irreducible unitary representation of $SL_2(\mathbf{R})$ has the strong mixing property (actually they formulated their result differently), Gelfand and Fomin gave a new proof of the results of Hedlund and Hopf. The Gelfand-Fomin argument went through several stages of generalization until today it can be used to show that nearly any measurable dynamical system coming from a homogeneous space will be ergodic, unless it fails to be for obvious reasons [**HoMo**, **Moor2**, **Zimm1**].

The influence between Lie theory and ergodic theory has been mutual. A particularly striking example of this was Margulis's use of ergodic theory in the proof of his Superrigidity Theorem [**Zimm1**], which was then reinterpreted as being a result in ergodic theory by Zimmer. A very recent example of this mutual interaction is the Margulis proof of the Oppenheim Conjecture [**Marg1**], followed quickly by Ratner's [**Ratn4**] broad generalization of the key ergodic-theoretic result underlying his proof.

As techniques for establishing ergodicity of dynamical systems became established, workers in ergodic theory refined their investigations. One kind of finer question frequently raised is loosely referred to as "unique ergodicity." It is a kind of inverse to the standard ergodic problem. Suppose (X, T) is a dynamical system and X is a locally compact Hausdorff space. Then X will support many probability measures, and one can ask for a description of all T-invariant probability measures on X. These will form a closed convex set in the unit ball of the space of all measures, and the ergodic ones are essentially the extreme points. Hence ergodicity figures in this question too. There is also clearly a connection between knowing all possible ergodic measures and knowing the closures of orbits.

Ratner [**Ratn1–4**] has more or less definitively answered these questions for the action of unipotent groups on homogeneous spaces. For purposes of explaining this, we will call an element g of a Lie group G *unipotent in* G if all eigenvalues of $\mathrm{Ad}\, g$ acting on the Lie algebra of G are equal to 1 (sometimes this is called Ad-unipotence). We will call a subgroup $H \subseteq G$ *unipotently generated* if the subset of elements of H which are unipotent in G generate H as group.

THEOREM 4.3.5. (a) *Let H be a unipotently generated subgroup of the Lie group G, and let $\Gamma \subseteq G$ be a lattice. For $x \in \Gamma\backslash G$, consider the closure* $\mathrm{Cl}(xH)$ *of the H orbit of x in $\Gamma\backslash G$. Then there is a subgroup $\widetilde{H}_x \subseteq G$, containing H, such that*

(i) $\mathrm{Cl}(xH) = x\widetilde{H}_x$,
(ii) $x^{-1}\Gamma x$ *is a lattice in* $\widetilde{H}_x$.

(b) *Let μ be an H-invariant probability measure on $\Gamma\backslash G$. Then there is a subgroup $\widetilde{H}_\mu \subseteq G$, containing H, and a point $x \in \Gamma\backslash G$ such that*

(i) $x^{-1}\Gamma x \cap \widetilde{H}_\mu$ *is a lattice in* $\widetilde{H}_\mu$,
(ii) μ *is the invariant probability measure on* $((x^{-1}\Gamma x)\cap\widetilde{H}_\mu)\backslash\widetilde{H}_\mu \simeq x\widetilde{H}_\mu$.

In particular, these results hold if H is a one-parameter subgroup of unipotent elements. It follows from this result and the compactness criterion [**BoHC**, **MoTa**] for arithmetic lattices, that if H is a one-parameter unipotent subgroup of $G = \mathrm{SL}_2(\mathbf{R})$, and if Γ is an arithmetic lattice with $\Gamma\backslash G$ compact, then

(i) *every* H-orbit is dense,
(ii) the only H-invariant measure is the standard measure on $\Gamma\backslash G$.

Such extraordinary rigidity of behavior is in striking contrast to the situation when H consists of noncompact semisimple elements, e.g., elements from the split Cartan subgroup A of the Iwasawa decomposition (cf. §A.2.3.3). For such H, Dani [**Dani3**] has shown the resulting transformations on $\Gamma\backslash G$ are isomorphic to Bernoulli shifts—and therefore determined up to isomorphism by their entropy. The rigidity of unipotent flows comes from the long term coherence of their trajectories, which diverge from one another slowly. In current parlance, unipotent flows are not chaotic [**Deva**]; whereas flows defined by semisimple elements have rapidly diverging trajectories and are chaotic (many of them are Anosov flows [**AbMa**, **ArAv**]). The coherence of trajectories, which Ratner terms "Property H," is illustrated by the following calculation comparing the "LU" with the "UL" decompositions (cf. §1.1) for $\mathrm{SL}_2(\mathbf{R})$:

$$\begin{bmatrix} a & 0 \\ z & a^{-1} \end{bmatrix} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b & 0 \\ 0 & b^{-1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ w & 1 \end{bmatrix}, \tag{4.3.6a}$$

where

$$s = \frac{a^2 t}{1 + azt}, \qquad b = \frac{a}{azt + 1}, \qquad w = \frac{az}{azt + 1}. \tag{4.3.6b}$$

We interpret these formulas as follows. We imagine we flow by right multiplications of the group $N^+ = \{[\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}]\}$ (t connoting time). From the N^+-trajectory through some point x_0, we move to a point x_1 by the small motion $[\begin{smallmatrix} a & 0 \\ z & a^{-1} \end{smallmatrix}]$ in a direction transverse to the flow lines, and then flow along the N^+-trajectory through x_1. This is the procedure indicated by the left side of equation (4.3.6a). The right-hand side tells us how to track the flow on the nearby N^+ trajectory by moving along the original trajectory (through x_0), at not quite a steady pace, specified by s, then moving transversely to the other trajectory, first by $[\begin{smallmatrix} b & 0 \\ 0 & b^{-1} \end{smallmatrix}]$ (which preserves the set of all N^+-trajectories), then by $[\begin{smallmatrix} 1 & 0 \\ w & 1 \end{smallmatrix}]$. The point to note is that, as long as we do not flow too far (precisely, as long as $1 + azt$ remains near 1, or at least, away from zero) we do not have to move very far to get from our original trajectory to the adjacent one. Further, the nearer the perturbed trajectory, the longer we can track it easily: if we divide z by 2, then the time until $b - 1$ or w exceed some prespecified limit at least doubles. If we were dealing with the rapidly diverging trajectories defined by semisimple elements, halving the original perturbation would only delay the time to divergence by a fixed increment, rather than doubling it.

Now let us describe a consequence of Theorem 4.3.5, the proof of a conjecture of Oppenheim **[Oppe]**, which was proved by Margulis **[Marg]** (see also **[DaMa2]**) on the basis of the relevant special case of Theorem 4.3.5. Consider a quadratic form

$$Q(x) = \sum_{j,k=1}^{n} b_{jk} x_j x_k \tag{4.3.7}$$

on $\mathbf{R}^n$. If Q is rational in the sense that its coefficients b_{jk} are rational numbers, then the values of Q on $\mathbf{Z}^n$ live in $\frac{1}{r}\mathbf{Z}$ for an appropriate denominator r; in particular the values form a discrete set. This remains true if Q is projectively rational in the sense that the ratios b_{jk}/b_{lm} are in $\mathbf{Q}$. It is also true if Q is definite. The conjecture of Oppenheim said essentially that these were the only conditions under which Q will have discrete values.

THEOREM 4.3.8. *Suppose $n \geq 3$, and that the form Q of formula (4.3.7) is indefinite and not projectively rational. Then the set of values $\{Q(z) : z \in \mathbf{Z}^n\}$ is dense in $\mathbf{R}$.*

To prove this using Theorem 4.3.5, observe first that it suffices to treat the case $n = 3$. Let $G = \mathrm{SL}_3(\mathbf{R})$, and let $H = \mathrm{SO}(Q)$ be the determinant one isometry group of Q. Consider the transforms $h(\mathbf{Z}^3)$ of the lattice $\mathbf{Z}^3 \subseteq \mathbf{R}^3$ by elements of H. This will be a collection of lattices inside the set of all lattices $L \subseteq \mathbf{R}^3$ such that the volume of $\mathbf{R}^n/L$ (measured with the push-down of Lebesgue measure) is 1, which set is identifiable with $\mathrm{SL}_3(\mathbf{Z})\backslash \mathrm{SL}_3(\mathbf{R})$. If the set $\{h(\mathbf{Z}^3), h \in H\}$ is dense in the space of lattices, then in particular the set of points $\{h(z) : h \in h, z \in \mathbf{Z}^3\}$ is dense in $\mathbf{R}^3$. On the other hand, since H preserves Q, the values of Q on $h(\mathbf{Z}^3)$ are the same as the values on $\mathbf{Z}^3$. Hence $Q(\mathbf{Z}^3)$ must be dense in $\mathbf{R}$.

So suppose $\{h(\mathbf{Z}^3)\}$ is not dense in the space of lattices. Since the form Q is indefinite, the group $\mathrm{SO}(Q)$, or at least its identity component, which has index 2 in the full group, is generated by unipotent elements. Hence Theorem 4.3.5 gives us a group $\widetilde{H}$, $H \subseteq \widetilde{H} \subseteq G$, such that the closure of $H(\mathbf{Z}^3)$ is $\widetilde{H}(\mathbf{Z}^3)$. But there are no groups between H and G (check the adjoint action of H on Lie $G/$Lie H: it is irreducible). Hence either $\widetilde{H} = G$ or $\widetilde{H} = H$. The first possibility amounts to the denseness of $H(\mathbf{Z}^3)$, which has already been rejected. Hence $H(\mathbf{Z}^3)$ is closed, and $H \cap \mathrm{SL}_3(\mathbf{Z})$ is a lattice in H. The Borel Density Theorem **[Zimm1]** then implies that $H \cap \mathrm{SL}_3(\mathbf{Z})$ is Zariski-dense in H, and from this Theorem 4.3.8 follows directly; for if we conjugate everything by an automorphism of $\mathbf{C}$ over $\mathbf{Q}$, the group $H \cap \mathrm{SL}_3(\mathbf{Z})$ will remain fixed, since it consists of matrices with rational (indeed, integer) entries. Hence its Zariski closure H remains fixed. Hence Q, which is determined up to multiples by H, remains fixed up to multiples. Hence the ratios b_{jk}/b_{lm} remain fixed, hence are in $\mathbf{Q}$.

Our second topic in dynamical systems will be essentially completely opposite from ergodic theory—completely integrable systems. Stimulated by discoveries of several examples of such systems, this topic was extremely active in the decade around 1980. Probably greatest interest attached to several infinite-dimensional systems associated to the "inverse scattering problem" **[BeCo, BeDT, TrPo]** for second-order ordinary differential operators, in particular the Korteweg-DeVries equation **[Adle2, GGKM, BeDT]**. However, some finite-dimensional systems were found which had strong analogies with the infinite-dimensional ones, and these, especially the Toda lattice **[GoWa, Kost6, Syme]**, also received considerable attention. Both the infinite-dimensional and the finite-dimensional systems were found to have close connections with Lie groups. We will discuss the (nonperiodic) Toda lattice because it is quite accessible and because it highlights the geometry of Lie groups, especially the relations between the various standard decompositions (Cartan, Iwasawa, Bruhat, cf. §A.2.3). Our account is largely based on **[GoWa]** and **[DLNT]**.

We should recall some features of Hamiltonian dynamics. We refer to §3.2 or **[AbMa, GuSt1]** for background about symplectic manifolds. Let M be a symplectic manifold, let q be a function on M, and let $\alpha^{-1}(dq)$ be the Hamiltonian vector field associated to f (cf. formula (3.2.1.6)). Let q_2 be another function on M. From our discussion in §3.2, we know that the conditions

$$\alpha^{-1}(dq)(q_2) = 0 \tag{4.3.9a}$$

and

$$[\alpha^{-1}(dq), \alpha^{-1}(dq_2)] = 0 \tag{4.3.9b}$$

are equivalent. However, the geometric interpretations of these two equations are quite different. Equation (4.3.9a) means that q_2 will be invariant under the flow generated by $\alpha^{-1}(dq)$. In other words, the flow defined by $\alpha^{-1}(dq)$ takes place in the level sets of q_2. Thus an equation of type (4.3.9a) helps us locate the trajectories of the flow of $\alpha^{-1}(dq)$. To suggest these roles, q_2 is sometimes called a *conserved quantity*, or an *integral* for the flow of $\alpha^{-1}(dq)$. On the other hand, equation (4.3.9b) means that the flows associated to $\alpha^{-1}(dq)$ and $\alpha^{-1}(dq_2)$ will commute with each other, hence each permutes the trajectories of the other, or each defines a one-parameter group of symmetries of the other. Thus, if we have two functions satisfying equations (4.3.9), both their flows are contained in their simultaneous level sets, and preserve each other. The fact that for Hamiltonian flows a conserved quantity plays a dual role of symmetry, and vice versa, is a particularly enriching feature of Hamiltonian mechanics.

Suppose that, given $q = q_1$ as above, we can find several functions q_2, $q_3, \ldots, q_m$ such that any pair q_j, q_k, $1 \leq j, k \leq q_m$, satisfies the mutually equivalent conditions (4.3.9). Then all the flows generated by the $\alpha^{-1}(dq_j)$, which (under the obvious necessary condition of functional independence, viz., that the vector fields (equivalently the differentials dq_j) should be linearly independent) will fill out submanifolds of dimension m, will all simultaneously preserve the level sets of all the q_j, which (under the same assumption of functional independence) will have codimension m. Thus we must have $m \leq \dim M - m$, or $\dim M \geq 2m$. In the extreme case, when $\dim M = 2m$, the systems of trajectories of the $\alpha^{-1}(dq_j)$ will completely fill, at least locally, the joint level surfaces of the q_j, so that the whole situation is determined: the possible motions generated by the $\alpha^{-1}(dq_j)$ completely fill up the surfaces defined by constancy of the q_j, and vice versa. Such a system is called *completely integrable.*

In two dimensions, i.e., if $\dim M = 2$, all Hamiltonian systems are completely integrable, but in dimensions greater than two, complete integrability is very special, and the discovery of completely integrable systems is an interesting event. In the 1960s, the physicist Toda **[Toda1]** defined a dynamical system interpretable as a system of n points moving on the real line subject to attractive forces depending exponentially on the distance between the particles. Precisely, it is the Hamiltonian system in $\mathbf{R}^{2m}$, with its standard symplectic form, defined by the function (in symplectic coordinates x_j, y_j, $1 \leq j \leq n$)

$$\frac{1}{2}\sum_{j=1}^{n} y_j^2 + \sum_{j=1}^{n-1} \exp(x_j - x_{j+1}). \tag{4.3.10}$$

It is easy to check that the subspace of $\mathbf{R}^{2n}$ defined by

$$\sum_{j=1}^{n} y_j = 0 = \sum_{j=1}^{n} x_j$$

is invariant for the flow defined by the Hamiltonian (4.3.10). If on this subspace we make the change of variables

(4.3.11)
$$a_j = -\frac{y_j}{4}, \quad b_k = \frac{1}{2}\exp\frac{(x_k - x_{k+1})}{2}, \qquad 1 \le j \le n, \, 1 \le k \le n-1,$$

then we find the equations

$$\frac{da_j}{dt} = (b_{j-1}^2 - b_j^2), \qquad \frac{db_k}{dt} = 2b_k(a_k - a_{k+1}). \tag{4.3.12}$$

We may interpret the equations (4.3.12) in terms of matrices in two different ways.

Flaschka **[Flas]** and Moser **[Mose]** interpreted equations (4.3.12) as follows. Let M^+ be the tridiagonal symmetric matrix

$$M^+ = \frac{1}{2}\begin{bmatrix} 2a_1 & b_1 & 0 & \cdots & 0 \\ b_1 & 2a_2 & b_2 & & 0 \\ 0 & & \ddots & & 0 \\ \vdots & & & & b_{n-1} \\ 0 & \cdots & & b_{n-1} & 2a_n \end{bmatrix} \tag{4.3.13}$$

and let M^- be the skew-symmetric matrix

$$M^- = \frac{1}{2}\begin{bmatrix} 0 & -b_1 & 0 & \cdots & 0 \\ b_1 & 0 & -b_2 & & \vdots \\ & b_2 & 0 & & \\ \vdots & & & \ddots & -b_{n-1} \\ 0 & \cdots & & b_{n-1} & 0 \end{bmatrix}. \tag{4.3.14}$$

Then the system of equations (4.3.12) can be checked to be equivalent to the matrix equation

$$\frac{dM^+}{dt} = 2[M^-, M^+]. \tag{4.3.15}$$

Adler **[Adle1]** and Kostant **[Kost6]** gave another interpretation of the system (4.3.12). Let $\mathbf{b}^0$ be the Lie algebra of upper triangular matrices of trace zero and $\mathbf{n}^+$ the subalgebra of strictly upper triangular matrices. By means of the bilinear form on $n \times n$ matrices

$$X, Y \to \operatorname{tr}(XY), \qquad X, Y \in M_n(\mathbf{R}), \tag{4.3.16}$$

where tr denotes the trace of a matrix, we can identify the space $M_n(\mathbf{R})$ of real $n \times n$ matrices with its dual. In this identification, the traceless matrices

$M_r(\mathbf{R})^0 \simeq \mathbf{sl}_n(\mathbf{R})$ are self-dual, and the dual of $\mathbf{b}^{0+}$ is identified to $\mathbf{sl}_n(\mathbf{R})$ modulo its annihilator, which is $\mathbf{n}^+$. Thus

$$(\mathbf{b}^{0+})^* \simeq \mathbf{sl}_n(\mathbf{R})/\mathbf{n}^+ \simeq \mathbf{b}^{0-}, \tag{4.3.17}$$

where $\mathbf{b}^{0-}$ is the lower triangular matrices of trace zero. In $\mathbf{b}^{0-}$, consider the set of matrices

$$M = \begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 \\ b_1 & a_2 & 0 & & \\ 0 & b_2 & a_3 & & \\ & & & \ddots & \\ 0 & \cdots & 0 & b_{n+1} & a_n \end{bmatrix}, \qquad b_j \neq 0, \sum_{j=1}^{n} a_j = 0. \tag{4.3.18}$$

It is easy to check that under the coadjoint action $\mathrm{Ad}^* B^{0+}$ (where $B^{0+} = \exp \mathbf{b}^{0+}$) on $\mathbf{b}^{0-} \simeq (\mathbf{b}^{0+})^*$, the matrices (4.3.18) form a single orbit, which we will refer to as the *Toda orbit*. One can also verify that the invariant symplectic form (cf. §3.2) on the Toda orbit is, up to multiples,

$$\omega = \sum_{1 \leq k \leq j \leq n-1} \frac{db_j \wedge da_k}{b_j} = \sum_{j=1}^{n-1} \frac{db_j \wedge dh_j}{b_j}, \tag{4.3.19}$$

where $h_j = \sum_{k=1}^{j} a_k$. If on this orbit one takes the Hamiltonian function

$$\frac{1}{2}\sum_{j=1}^{n-1} b_j^2 + \sum_{k=1}^{n} a_k^2 = \frac{1}{4}\operatorname{tr}(M + M^t)^2, \tag{4.3.20}$$

then the Hamiltonian flow defined by (4.3.20) with respect to the symplectic form (4.3.19) is again the system (4.3.12). In equation (4.3.20), M^t denotes the transpose of the matrix M.

Thus we have two Lie-theoretic interpretations of the Toda lattice: one as a Hamiltonian flow on a coadjoint orbit for B^{0+}, with respect to a quadratic Hamiltonian related to the embedding $\mathbf{b}^{0+} \subseteq \mathbf{sl}_n(\mathbf{R})$, the other as a flow on the tridiagonal symmetric matrices, defined by commutator with a skew-symmetric matrix. The two interpretations are related by the equations

$$M^+ = \tfrac{1}{2}(M + M^t), \qquad M^- = \tfrac{1}{2}(M - M^t) \tag{4.3.21}$$

with M as in equation (4.3.18) and M^+ and M^- as in (4.3.13) and (4.3.14).

The form (4.3.15) of the Toda lattice equations, involving the commutator of a matrix with a skew-symmetric matrix depending on it, is known as *Lax form*. Its significance is that, since commutator is the infinitesimal version of conjugation, it immediately implies that all matrices in the trajectory through M^+ will be conjugate to M^+ (and by an orthogonal matrix). Hence the flow will be *isospectral*; the eigenvalues of M^+ will be constant on each trajectory of the flow. If these eigenvalues (or, more properly, symmetric functions of them) are pulled back to the Toda orbit (4.3.18) via the map (4.3.21), they define $n-1$ integrals of the Toda flow. Since the Toda orbit has dimension $2(n-1)$, we see that if these integrals of the Toda flow Poisson commute with each other (i.e., if their associated Hamiltonian vector fields commute), then the Toda flow will be completely integrable, and the orbits of all the flows together will fill out the tridiagonal matrices of given eigenvalues. This is indeed the case, as we shall show.

Involved in this story are several actions related to one another by the geometry of $G = \mathrm{SL}_n(\mathbf{R})$. As above, let B^{0+} be the connected group whose Lie algebra is $\mathbf{b}^{0+}$. Let $K = \mathrm{SO}_n(\mathbf{R})$—this is a maximal compact subgroup of $\mathrm{SL}_n(\mathbf{R})$. We have the Iwasawa, or for $\mathrm{SL}_n(\mathbf{R})$, the Gram-Schmidt decomposition (cf. §A.2.3)

$$\mathrm{SL}_n(\mathbf{R}) = B^{0+}K, \qquad K \cap B^{0+} = 1. \tag{4.3.22}$$

Thus we have an identification

$$B^{0+}\backslash \mathrm{SL}_n(\mathbf{R}) \simeq K \tag{4.3.23}$$

so that the action of K on itself by right translation extends to an action of all of $\mathrm{SL}_n(\mathbf{R})$. Understanding the Toda lattice involves looking at this action from several points of view.

The first point of view is to express the $\mathrm{SL}_n(\mathbf{R})$ action pointwise in terms of the K-action. Let

$$g = b(g)k(g), \qquad g \in \mathrm{SL}_n(\mathbf{R}), \tag{4.3.24}$$

be the decomposition of an element g in $\mathrm{SL}_n(\mathbf{R})$ according to (4.3.22): thus $b(g) \in B^{0+}$ and $k(g) \in K$. We clearly have

$$b(b_1 g k_1) = b_1 b(g), \qquad k(b_1 g k_1) = k(g)k_1 \tag{4.3.25}$$

for $g \in \mathrm{SL}_n(\mathbf{R})$, $b_1 \in B^{o+}$, $k_1 \in K$. We also observe that $g \to k(g)$ is the mapping which implements the identification (4.3.23). We may compute

$$k(k_1 g) = k((k_1 g k_1^{-1})k_1) = k(k_1 g k_1^{-1})k_1. \tag{4.3.26}$$

We can also express this infinitesimally. If $g = \exp tX$, for $X \in \mathbf{sl}_n(\mathbf{R})$, then we can find $\kappa(X, k_1)$ such that

$$\frac{d}{dt}(k_1^{-1}k(k_1 \exp tX))|_{t=0} = \kappa(X, k_1).$$

To express κ, we need the decomposition

$$\mathbf{sl}_n(\mathbf{R}) \simeq \mathbf{so}_n \oplus \mathbf{b}^{0+} \simeq \mathbf{k} \oplus \mathbf{b}^{0+}, \tag{4.3.27}$$

which is the infinitesimal version of decomposition (4.3.22). Let

$$p_k : \mathbf{sl}_n(\mathbf{R}) \to \mathbf{k}, \qquad p_b : \mathbf{sl}_n(\mathbf{R}) \to \mathbf{b}^{0+} \tag{4.3.28}$$

be the projection maps associated with equation (4.3.27). In terms of these maps, by plugging $g = \exp tX$ into formula (4.3.26) and differentiating, we can compute

$$\kappa(X, k_1) = \mathrm{Ad}\, k_1^{-1}(p_k(\mathrm{Ad}\ k_1(X))). \tag{4.3.29}$$

Now consider the Cartan decomposition (cf. §A.2.3.1)

$$\begin{aligned} (4.3.30) \qquad & \mathbf{sl}_n(\mathbf{R}) \simeq \mathbf{k} \oplus \mathbf{p}, \\ & M = M^+ + M^- = \tfrac{1}{2}(M + M^t) + \tfrac{1}{2}(M - M^t) \end{aligned}$$

of a matrix into its symmetric and skew-symmetric parts. The summands $\mathbf{k}$ and $\mathbf{p}$ are invariant under $\mathrm{Ad}\, K$. In particular, we have an action of K on $\mathbf{p}$, the space of symmetric matrices. The basic theorem on diagonalizing symmetric matrices tells us that

$$(4.3.31) \qquad \mathrm{Ad}\, K(\mathbf{a}^0) = \mathbf{p}.$$

Furthermore, if X in $\mathbf{a}^0$ is regular, in the sense that all the eigenvalues of X are distinct (so that no roots of $\mathbf{a}^0$ in $\mathbf{sl}_n(\mathbf{R})$, which are differences of eigenvalues, vanish on X), then the stabilizer of X in K is discrete, so that the map $k \to \mathrm{Ad}\, k(X)$ is almost injective, i.e., it defines a covering by K of the $\mathrm{Ad}\, K$ orbit of X.

The decomposition (4.3.30) is orthogonal with respect to the inner product (4.3.16), hence that inner product restricts nondegenerately to $\mathbf{p}$. (In fact, it is positive definite.) Given a function f on $\mathbf{p}$, we can form its gradient ∇f in the standard way **[CoSe; Lang2**, p.186], by the formula

$$(4.3.32) \qquad \mathrm{tr}(\nabla f(y) z) = df(y)(z) = \partial_z(f)(y), \qquad y, z \in \mathbf{p},$$

where $df(y) \in \mathbf{p}^*$ is the usual differential of f at y, and $\partial_z f$ indicates the directional derivative in the direction z. Using ∇f and the projection p_k (cf. (4.3.28)) to $\mathbf{k}$ along $\mathbf{b}^{0+}$, we can define a Lax form dynamical system on $\mathbf{p}$ by the differential equation

$$(4.3.33) \qquad \frac{dy}{dt} = [p_k(\nabla f(y)), y].$$

This will give a flow which preserves each $\mathrm{Ad}\, K$-orbit in $\mathbf{p}$, since the tangent vector at each point y is by definition in $\mathrm{ad}\mathbf{k}(y)$, which is the tangent space to $\mathrm{Ad}\, K(y)$ at y. Thus for each regular y, we get by transport of structure a flow on K. This is explicitly defined as follows. Fix a regular y_0, and represent the general y in $\mathrm{Ad}\, K(y_0)$ in the form $y = \mathrm{Ad}\, k_1(y_0)$. Then equation (4.3.33) reads

$$\begin{aligned} \frac{d}{dt}(\mathrm{Ad}\, k_1(y_0)) &= [p_k \nabla f(\mathrm{Ad}\, k_1(y_0)), \mathrm{Ad} k_1(y_0)] \\ &= \mathrm{Ad}\, k_1 [\mathrm{Ad}\, k_1^{-1}(\nabla f(\mathrm{Ad}\, k_1(y_0)), y_0] \end{aligned}$$

or

$$(4.3.34) \qquad (\mathrm{Ad}\, k_1)^{-1} \frac{d}{dt}(\mathrm{Ad}\, k_1(y_0)) = \mathrm{ad}(\mathrm{Ad}\, k_1^{-1}(p_k(\nabla f(\mathrm{Ad}\, k_1(y_0)))(y_0).$$

Now suppose that the function f is invariant under $\mathrm{Ad}\, K$. From the chain rule

$$d(f \circ \mathrm{Ad}\, k) = ((df) \circ \mathrm{Ad}\, k)\mathrm{Ad}\, k, \qquad k \in K,$$

we deduce from the invariance of f $(f = f \circ \mathrm{Ad}\, k)$ and the definition (4.3.32) of ∇f, that

$$\nabla f(\mathrm{Ad}\, k(y)) = \mathrm{Ad}\, k(\nabla f(y)), \qquad k \in K, \; y \in \mathbf{p}. \tag{4.3.35}$$

If we plug this into equation (4.3.34), and suppress the y_0, writing $\nabla f(y_0) = X_0$, we obtain the following equation for k_1:

$$k_1^{-1}\frac{d}{dt}k_1 = \mathrm{Ad}\, k_1^{-1}(p_k(\mathrm{Ad}\, k_1(X_0))). \tag{4.3.36}$$

If we compare this with equation (4.3.29) for the right action of $\mathrm{SL}_n(\mathbf{R})$ on $B^{0+}\backslash \mathrm{SL}_n(\mathbf{R})$ we obtain the following result.

PROPOSITION 4.3.37. *Let f be an $\mathrm{Ad}\, K$-invariant function on $\mathbf{p}$. Then the flow on $\mathbf{p}$ defined by the Lax-type equations* (4.3.33) *is equivalent on each $\mathrm{Ad}\, K$-orbit $\mathrm{Ad}\, K(y_0)$ to a flow induced by right translations by a one parameter group of $\mathrm{SL}_n(\mathbf{R})$ acting on $B^{0+}\backslash \mathrm{SL}_n(\mathbf{R}) \simeq K$. Precisely, under the covering $K \to \mathrm{Ad}\, K(y_0)$ by the map $k_1 \to \mathrm{Ad}\, k_1(y_0)$, the flow* (4.3.33) *on $\mathrm{Ad}\, K(y_0)$ is identified to right translation on $B^{0+}\backslash \mathrm{SL}_n(\mathbf{R})$ by $\exp(t \nabla f(y_0))$.*

To tighten the connection made by Proposition 4.3.37, we need another observation about ∇f for $\mathrm{Ad}\, K$-invariant f. For regular $y_0 \in \mathbf{p}$, let $\mathbf{c}(y_0)$ be the centralizer of y_0 in $\mathbf{p}$. Thus if $y_0 = \mathrm{Ad}\, k_0(a_0)$, for $a_0 \in \mathbf{a}^0$, we have $\mathbf{c}(y_0) = \mathrm{Ad}\, k_0(\mathbf{a}^0)$. We have

$$\nabla f(y_0) \in \mathbf{c}(y_0). \tag{4.3.38}$$

Indeed, we know $\nabla f(y_0)$ is orthogonal to the level set of f through y_0. Since f is $\mathrm{Ad}\, K$-invariant, the level set includes $\mathrm{Ad}\, K(y_0)$, and the tangent space at y_0 to $\mathrm{Ad}\, K(y_0)$ is $\mathrm{ad}\, \mathbf{k}(y_0)$. Since we have assumed y_0 to be regular, this equals $\mathbf{c}(y_0)^{\perp}$, the orthogonal complement to $\mathbf{c}(y_0)$ in $\mathbf{p}$, as can easily be seen in the case $y_0 \in \mathbf{a}^0$. Since all elements of $\mathbf{c}(y_0)$ commute with one another, combining Proposition 4.3.37 with observation (4.3.38) yields the following conclusion.

COROLLARY 4.3.39. *All the flows on $\mathbf{p}$, associated to the Lax-type equations* (4.3.33) *for f which are $\mathrm{Ad}\, K$-invariant, commute with each other.*

To complete the picture, it remains to relate the Lax-type equations to Hamiltonian systems on coadjoint orbits. Here again the relation (4.3.38) is a key ingredient. It implies that, again taking f to be $\mathrm{Ad}\, K$-invariant,

$$[\nabla f(y), y] = 0, \qquad y \in \mathbf{p}. \tag{4.3.40}$$

Take x, y in $\mathbf{p}$, and decompose x as in equation (4.3.27). Then if x, y commute, we have

$$[p_k(x), y] = -[p_b(x), y], \qquad x, y \in \mathbf{p};\; [x, y] = 0. \tag{4.3.41}$$

Let f be an $\mathrm{Ad}\, K$-invariant function on $\mathbf{p}$. Extend f to a function on all $\mathbf{g}$ by letting f be constant on cosets of $\mathbf{k}$. Then $\nabla f(y)$, $y \in \mathbf{p}$, is the same as when f was considered as a function on $\mathbf{p}$. Thus from (4.3.40) and (4.3.41) we deduce

$$[p_k(\nabla f(y)), y] = -[p_b(\nabla f(y)), y]. \tag{4.3.42}$$

There is a natural projection

$$\alpha : \mathbf{g} \to (\mathbf{b}^{0+})^* \tag{4.3.43}$$

induced by the bilinear form (4.3.16). The restriction of α to $\mathbf{p}$ is a linear isomorphism of $\mathbf{p}$ onto $(\mathbf{b}^{0+})^*$. Since α is $\mathrm{Ad}\, B^{0+}$-equivariant, we have

$$\alpha(\mathrm{ad}\, p_b(\nabla f(y))(y)) = \mathrm{ad}\, p_b(\nabla f(y))\alpha(y). \tag{4.3.44}$$

The function f on $\mathbf{p}$ pushes down via α to define a function $H_f = f \circ \alpha^{-1}$ on $\mathbf{b}^{0+}$. Checking through the various identifications and using equations (4.3.42) and (4.3.44) we can verify **[GoWa, Syme]**

PROPOSITION 4.3.45. *Let f be an $\mathrm{Ad}\, K$-invariant function on $\mathbf{p}$, and let $H_f = f \circ \alpha^{-1}$ be the push-forward of f to $(\mathbf{b}^{0+})^*$ via α. Then the Lax-type system* (4.3.33), *pushed forward to $(\mathbf{b}^{0+})^*$ by α, becomes the Hamiltonian system defined by $-H_f$ on each B^{0+} coadjoint orbit in $\mathbf{b}^{0+}$.*

If we now combine Proposition 4.3.45 with Corollary 4.3.39, we find the $n-1$ Hamiltonians corresponding to $\mathrm{trace}(\alpha^{-1}(M))^k$, for $k = 2, 3, \ldots, n$, form a Poisson commutative family on any $\mathrm{Ad}^* B^{0+}$-orbit in $(\mathbf{b}^{0+})^*$. In particular, since the Toda orbit (4.3.18) has dimension $2(n-1)$ they define on it (assuming they are functionally independent, which is easy to check) a completely integrable family. Thus the Toda lattice is completely integrable.

REMARKS. (a) The argument above shows that the $H_f = f \circ \alpha^{-1}$ form a Poisson commutative family on any $\mathrm{Ad}^* B^{0+}$-orbit in $(\mathbf{b}^{0+})^*$. However, only for special orbits which have dimension $2(n-1)$ or less could we hope to conclude complete integrability from this fact. In **[GoWa]** other orbits where this scheme gives complete integrability are described. On the other hand, in **[DLNT]** more Poisson commuting functions are found on $(\mathbf{b}^{0+})^*$, enough to provide complete integrability for the Toda flow on a generic $\mathrm{Ad}^* B^{0+}$-orbit in $(\mathbf{b}^{0+})^*$.

(b) Systems analogous to the Toda lattice can be constructed on any split real semisimple group **[Kost6, Syme1]**.

(c) A number of other completely integrable systems on coadjoint orbits have been found. For example, by considering a nested chain $\mathbf{u}_1 \subseteq \mathbf{u}_2 \subseteq \mathbf{u}_3 \subseteq \cdots \subseteq \mathbf{u}_n$ of unitary Lie algebras Thimm(**[Thim]** (see also **[GuSt5]**) was able to construct completely integrable systems on coadjoint orbits in $\mathbf{u}_n^*$. Included among these are the geodesic flows on Grassmannians (cf. §1.4). Similar considerations apply to orthogonal Lie algebras.

(d) Using the Bruhat decomposition (cf. §A.2.3.3), one can give a very clear description of the dynamics of right translation on $B^{0+}\backslash \mathrm{SL}_n(\mathbf{R})$ by

$\exp tX$, $X \in \mathbf{p}$. It behaves very simply: it is essentially a gradient flow, with nondegenerate, isolated fixed points, parametrized by the Weyl group. In fact, the Bruhat decomposition arises as the Morse decomposition **[Miln2]** associated to this flow. Similar facts can be shown to hold for the holomorphic action of $\mathbf{C}^{\times}$ on projective varieties **[Carr]**. (Of course $B^{0+}\backslash \mathrm{SL}_n(\mathbf{R})$ is not a complex variety, but if we had been working with $\mathrm{SL}_n(\mathbf{C})$ rather than $\mathrm{SL}_n(\mathbf{R})$, we would have been dealing with the complex flag manifold, which is a complex variety.) The geometric analysis allows one to describe quite precisely the asymptotics of Toda trajectories.

(e) The Toda lattice can be solved quite explicitly **[Kost6, Syme1, GoWa]**. It turns out to be closely related to the famous "QR algorithm" for diagonalizing matrices. It is essentially a continuous-time version of this algorithm **[Syme2, DeNT, GoWa]**.

(f) It should be noted that Proposition 4.3.45 is a generalization of the results of the explicit calculations (4.3.10) through (4.3.21).

Appendix 1: Basic concepts of representation theory. An account of representation theory must begin with some basic definitions and some remarks about certain technical issues. These latter can be somewhat off-putting, but if openly acknowledged, their negative effects can be minimized. The reader is advised to skim this section, and refer to it as necessary. For greater detail on this basic material, we refer to **[FeDo, Gaal, Kiri, Lang1]**, etc.

A.1.1. Let G be a Lie group. A *representation* ρ of G on a vector space V is a homomorphism of G into the group of invertible linear transformations of V:

$$\rho : G \to \mathrm{GL}(V). \tag{A.1.1.1}$$

To be complete, in referring to a representation, we should specify both ρ and V, but often we will only specify ρ, letting V be understood implicitly; or we may just specify V and let ρ be implicit, in which case we call V a *G-module.*

A.1.2. Very often V will be infinite dimensional, and then usually it is equipped with a topology. Although the case of greatest general interest is when V is a Hilbert space, sometimes it is a Banach space, and it is not really possible to avoid considering situations when V is only locally convex. Whatever the topology of V, one wants to put a continuity condition on ρ. The correct one is that ρ should be *strongly continuous*:

(A.1.2.1) The map $g \to \rho(g)v$, from G to V, should be continuous for all $v \in V$.

REMARK. If V is a Banach space, one might be tempted to think the map $g \to \rho(g)$ should be continuous with respect to the norm topology on the operators on V. But this condition is far too restrictive, and hardly ever holds when V is infinite dimensional.