4. For the 'satiably curious, here it is: A (linear) algebraic subgroup of $\mathrm{GL}_n(\mathbf{C})$ is a subgroup which is also an algebraic subvariety, i.e., the set of zeros of a collection of polynomials in X_{ij}, the coordinate functions on M_n, and in $\det^{-1}$, the reciprocal of the determinant function. Thus $\mathrm{SL}_n(\mathbf{C})$, the special linear group, is defined by the equation $\det g = 1$; and $\mathrm{O}_n(\mathbf{C})$, the complex orthogonal group, is defined by the equation $g^t g = 1$, which can be regarded as the collection of n^2 scalar equations

$$\sum_{j=1}^{n} g_{ji} g_{jk} = \delta_{ik} \quad \text{for } 1 \leq i, k \leq n.$$

Let $G \subseteq \mathrm{GL}_n(\mathbf{C})$ be an algebraic subgroup, and let I_G be the ideal of polynomials vanishing on G. If we can find a generating set $\{P_j\}_{j=1}^{m}$ of polynomials for I_G such that the coefficients of the P_j are real, we say G is *defined over* $\mathbf{R}$. If we can find such P_j with coefficients in $\mathbf{Q}$, we say G is *defined over* $\mathbf{Q}$. If G is defined over $\mathbf{R}$, then

$$G_{\mathbf{R}} = G \cap \mathrm{GL}_n(\mathbf{R})$$

is called the *real points* of G. Similarly for $\mathbf{Q}$.

If G is a linear algebraic group defined over $\mathbf{Q}$, then $G \cap \mathrm{GL}_n(\mathbf{Z}) = G_{\mathbf{R}} \cap \mathrm{GL}_n(\mathbf{Z})$ is called the subgroup of *integral points of* G. Two subgroups Γ_1, Γ_2 of a group G are called *commensurable* if $\Gamma_1 \cap \Gamma_2$ has finite index in both Γ_1 and Γ_2.

Let G_0 be a Lie group, and $\Gamma \subseteq G_0$ a discrete subgroup. We say Γ is *arithmetic* if there exist

(i) a linear algebraic group $G \subseteq \mathrm{GL}_n(\mathbf{C})$ defined over $\mathbf{Q}$, and
(ii) a homomorphism $\Psi: G_0 \to G_{\mathbf{R}}$, such that
(iii) $\ker \Psi$ is compact,
(iv) $\mathrm{im}\, \Psi$ is normal in $G_{\mathbf{R}}$,
(v) $G_{\mathbf{R}}/\mathrm{im}\, \Psi$ is compact, and, most importantly,
(vi) $\Psi^{-1}(G \cap \mathrm{GL}_n(\mathbf{Z}))/\ker \Psi$ is commensurable with $(\Gamma \cdot \ker \Psi)/\ker \Psi$.

5. It might be thought one should demand that G/Γ be compact; but this would exclude important examples including $\mathrm{SL}_n(\mathbf{Z})$. The relaxation of compactness to finiteness of covolume has been very fruitful.

6. A precise question of this nature was formulated by Selberg **[Selb1]**.

2. An outline of Lie theory.

2.1. The glue that binds Lie theory together is the notion of a one-parameter group and its infinitesimal generator. For expository purposes, we will consider the one-parameter group first, although this is a revisionist way of proceeding. This discussion can be found in many texts (e.g., **[Ster, Gilm, Pont, Hoch, Vara]**, etc.). But it is so basic, it seems necessary to include it.

Let M be a manifold. (You can just think of an open set in $\mathbf{R}^n$ if you wish.) The set $\mathrm{Diff}(M)$ of diffeomorphisms (smooth, smoothly invertible mappings) of M is a group with composition of mappings as the product. The most perspicuous way to think of a one-parameter group of diffeomorphisms of M is simply as a homomorphism

$$\begin{aligned}\phi: \mathbf{R} &\to \mathrm{Diff}(M) \\ t &\to \phi_t.\end{aligned} \tag{2.1.1}$$

However, technical considerations require certain smoothness conditions. These may most conveniently be formulated by requiring that the map

$$\Phi: \mathbf{R} \times M \to M, \qquad \Phi(t, m) = \phi_t(m), \quad t \in \mathbf{R}, m \in M, \tag{2.1.2}$$

be smooth.

Consider a one-parameter group of diffeomorphisms ϕ_t of M. Fix $m \in M$, and consider the curve

$$\gamma_m(t) = \phi_t(m). \tag{2.1.3}$$

Conditions (2.1.2) guarantee this is a smooth curve. It passes through m at $t = 0$. At that point (in time and space), the tangent vector to the curve is

$$v(m) = \gamma_m'(0) = \frac{d}{dt}\phi_t(m)|_{t=0}. \tag{2.1.4}$$

The map $v: m \to v(m)$ is a vector field on M—it assigns to each point m the tangent vector $v(m)$; condition (2.1.2) guarantees it is a smooth vector field.

REMARK. The intuitive geometric connection between the one-parameter group ϕ_t and the vector field v is most easily seen by taking, temporarily at least, M to be an open set in $\mathbf{R}^n$. Then equation (2.1.4) is equivalent to

$$\phi_t(m) = m + tv(m) + t^2\varepsilon(t, m),$$

where $\varepsilon(t)$ is a smooth function of t and m. Thus, for small times, the motion ϕ_t displaces m approximately by the vector $tv(m)$; this approximation becomes more accurate as $t \to 0$. Thus, if we allow ourselves the language of infinitesimals, we may say that after an infinitesimal time ε, the point m moves to $m + \varepsilon v(m)$. This was standard parlance in the 19th century, and the actual motion $m \to \phi_t(m)$ was thought of as being composed of a very large number of these very small motions $m \to m + \varepsilon v(m)$. (This intuition is justified rigorously by Euler's approximation scheme for solving O.D.E. **[GuNi, Zill]**, etc.) For this reason the vector field v was called the "infinitesimal generator" of the one-parameter group ϕ_t.

Now consider the tangent vectors to the curve γ_m at other times. We compute

$$\begin{aligned}\gamma_m'(s) &= \frac{d}{dt}\phi_t(m)|_{t=s} = \frac{d}{dt}((\phi_{t-s} \cdot \phi_s)(m))|_{t=s} \\ &= \frac{d}{dt}\phi_t(\phi_s(m))|_{t=0} = v(\phi_s(m)).\end{aligned} \tag{2.1.5}$$

Thus, the tangent vector to the curve γ_m at any point is the vector assigned by the vector field v of (2.1.4). In other words, the mapping $t \to \gamma_m(t)$ is a solution of the differential equation

$$\frac{d\gamma}{dt}(t) = v(\gamma(t)). \tag{2.1.6a}$$

Since $\gamma_m(0) = m$, we see that γ_m is the solution of equation (2.1.6a) with initial condition

$$\gamma_m(0) = m. \tag{2.1.6b}$$

The above reasoning applies for all times t and all m in M. Thus we see that having the one-parameter group ϕ_t gives us solutions, for all time t, and for all initial conditions m, of the differential equation (2.1.6a).

On the other hand, suppose we start with the differential equation (2.1.6a). The classical (nineteenth century—contemporaneous with Lie) Existence and Uniqueness Theorem for ordinary differential equations (see, for example, **[Ster, LoSt, BiRo]**) tells us that given any m, there is some $\varepsilon(m) > 0$ such that for $|t| < \varepsilon(m)$, there is a solution $\gamma_m(t)$ of equation (2.1.6a) with initial condition (2.1.6b). Moreover, this solution is unique. An easy extension of this basic result shows that in fact, for each m, there is some minimum number $t^-(m) < 0$ and some maximum number $t^+(m) > 0$ such that $\gamma_m(t)$ can be defined for $t^-(m) < t < t^+(m)$, and if $t^+(m) < \infty$, then as $t \to t^+(m)$, the curve $\gamma_m(t)$ "drops off the edge of the world" in the sense that $\gamma_m(t)$ has no limit points in M as $t \to t^+(m)$; similarly if $t^-(m) > -\infty$.

Consider a particular solution curve $\gamma_m(t)$ of (2.1.6a) with initial condition (2.1.6b). At time $t = s$, the curve passes through $\gamma_m(s)$. Consider the curve $\gamma_{m,s}$ obtained from γ_m by shifting the time variable:

$$\gamma_{m,s}(t) = \gamma_m(s+t). \tag{2.1.7}$$

Then we can compute

$$\frac{d}{dt}\gamma_{m,s}(t) = \frac{d}{dt}\gamma_m(s+t) = v(\gamma_m(s+t)) = v(\gamma_{m,s}(t)).$$

Thus $\gamma_{m,s}$ also satisfies the differential equation (2.1.6a); but $\gamma_{m,s}$ satisfies the initial condition $\gamma_{m,s}(0) = \gamma_m(s)$. From the uniqueness part of the Existence and Uniqueness Theorem, we conclude

$$\gamma_m(s+t) = \gamma_{m,s}(t) = \gamma_{\gamma_m(s)}(t). \tag{2.1.8}$$

Let us now assume that $\gamma_m(t)$ is defined for all t and all m.[1] Then for each t, we can define a map

$$\phi_t \colon M \to M \tag{2.1.9}$$

by the recipe

$$\phi_t(m) = \gamma_m(t).$$

With this change of notation, the relation (2.1.8) turns into

$$\phi_{t+s}(m) = \phi_t(\phi_s(m)),$$

that is, $\phi_{s+t} = \phi_s \circ \phi_t$. Since ϕ_0 is clearly the identity, we conclude ϕ_t and ϕ_{-t} are mutually inverse mappings. Hence each map ϕ_t is actually bijective and the map $t \to \phi_t$ is a homomorphism from $\mathbf{R}$ to the group of permutations of the points of M. Further, the Existence and Uniqueness Theorem has some standard complements concerning smoothness in the initial conditions which guarantee that the maps ϕ_t are smooth, hence diffeomorphisms, and even that the map ϕ defined as in (2.1.2) is smooth. Hence the ϕ_t form a one-parameter group of diffeomorphisms of M.

To sum up, we can enunciate the following correspondence principle, which amounts to a geometric/group theoretic interpretation of the Existence and Uniqueness Theorem for O.D.E.

(2.1.10) To every one-parameter group ϕ_t of diffeomorphisms of a manifold M is associated a vector field v, the "infinitesimal generator" of ϕ_t, by equation (2.1.4). Knowledge of ϕ_t is equivalent to the ability to solve, for all initial values m and all times t, the differential equations (2.1.6) associated to v.

Thus there is a one-to-one correspondence between one-parameter groups acting on M and certain vector fields on M, namely those for which the equations (2.1.6) can be integrated for all time. (If M is compact, this will be all vector fields.) In Lie's time, one was not so fastidious about the global requirement "for all m for all time," so one considered simply that there was a one-to-one correspondence between "one-parameter groups" (in the 19th century sense) and vector fields. Today, one achieves this one-to-one correspondence by replacing the one-parameter group by the one-parameter "local group" or "pseudo-group" **[GuSt2]**. This is not a group but a collection of mappings trying to fit together to be a group. It is the obvious formalization of what you get from the Existence and Uniqueness Theorem. It is a rather cumbersome technical notion which attempts, with only partial success, to restore to us the Eden we lost when we achieved awareness of global problems.

2.2. A fundamental class of examples of one-parameter groups is obtained by taking M simply to be a vector space, and requiring the ϕ_t to be linear transformations. Since the infinitesimal generator v of ϕ_t is obtained as a limit,

$$v(m) = \lim_{t\to 0} \frac{\phi_t(m) - m}{t} = \lim_{t\to 0}\left(\frac{\phi_t - 1}{t}\right)(m), \qquad m \in M, \tag{2.2.1}$$

we see that $m \to v(m)$ is likewise a linear transformation. Let us call it A. Thus

$$A = \lim_{t\to 0} \frac{\phi_t - 1}{t}, \tag{2.2.2}$$

where this limit is taken in the algebra $\mathrm{End}(M)$ of matrices on M. With this notation, the differential equations (2.1.6) specialize to

$$\frac{d\gamma}{dt} = A(\gamma). \tag{2.2.3}$$

Equation (2.2.3) will be recognized as a system of constant-coefficient homogeneous linear differential equations, such as occupy a large chunk of introductory courses on ordinary differential equations [**GuNi**, **BoDP**, **Zill**], and form the basis of linear system theory [**TiBo**, **ZaDe**].

We know how to solve equations (2.2.3) explicitly in terms of the matrix A. We form $\exp A$, the exponential of A, by means of the familiar power series for exp:

$$\exp A = 1 + A + \frac{A^2}{2} + \frac{A^3}{6} + \cdots + \frac{A^k}{k!} + \cdots. \tag{2.2.4}$$

Then termwise differentiation of the function $t \to \exp tA$ yields the equation

$$\frac{d}{dt}(\exp tA) = A \exp tA. \tag{2.2.5}$$

It follows that

$$\gamma_m(t) = \exp tA(m) \tag{2.2.6}$$

is a solution of (2.2.3) with initial value m at $t = 0$. Thus, in this special case, we can recover the one-parameter group from A defined by (2.2.2) by

$$\phi_t = \exp tA. \tag{2.2.7}$$

Because of formula (2.2.7), one often abuses terminology and calls A (rather than the associated vector field which assigns $A(v)$ to v) the infinitesimal generator of ϕ_t. Also, because of strong analogies between this special case and the general one-parameter group, one often refers to the procedure of constructing a one-parameter group from the vector field which is its infinitesimal generator as *exponentiating* the vector field.

2.3. We now have a grasp of the bedrock of Lie theory, the connection between a one-parameter group and its infinitesimal generator. The next step, which is the fundamental insight of Lie theory, is how to combine several one-parameter groups into a multi- (but finite!) parameter group—a Lie group. Roughly speaking, one finds that a Lie group is a very coherent collection of one-parameter subgroups.

To firm up ideas, we imagine, in analogy with formula (2.1.2), that we have a manifold M, and an auxiliary manifold G, which is parametrizing a group of diffeomorphisms of M. Thus we have a mapping

$$\Phi\colon G \times M \to M, \tag{2.3.1}$$

which we take to be smooth, such that for each $g \in G$ the map

$$\phi_g\colon M \to M, \qquad \phi_g(m) = \Phi(g, m), \tag{2.3.2}$$

is a diffeomorphism of M, and such that the maps ϕ_g form a group; the composition of two of them is a third one, the identity is one, the inverse of one is one, etc. An important example of such a G occurs when M is a vector space, and $G = \mathrm{GL}(M)$ is the group of all invertible linear transformations of M.

Inside the group G there will be various one-parameter subgroups, to each of which corresponds a unique infinitesimal generator. A priori these infinitesimal generators are just a set of vector fields. Let us call this set $\mathrm{Lie}(G)$. The magic comes in realizing that in fact this seemingly rather unwieldly object, the collection of infinitesimal generators of one-parameter subgroups of G, has a very precise structure: it is, first, a real vector space; and in addition, it has defined on it a skew-symmetric product—the Lie bracket. (The modern approach to Lie groups, which defines $\mathrm{Lie}(G)$ as the space of left-invariant vector fields on G, makes these facts virtually automatic. It is commendable in its efficiency, but it takes a lot of the wonder out of the story.)

The piece of algebraic structure on $\mathrm{Lie}(G)$ that is easiest to understand is scalar multiplication. If $t \to \phi_t$ is a one-parameter group of diffeomorphisms, with infinitesimal generator v, then $t \to \phi_{st}$, $s \in \mathbf{R}$, is obviously also a one-parameter group of diffeomorphisms, and its infinitesimal generator is easily checked to be sv. Thus the set $\mathrm{Lie}(G)$ is closed under multiplication by scalars.

The next observation is that the infinitesimal analog of multiplication of one-parameter groups is simply addition of vector fields. This is easily seen by the following formal, purely local computation. Let ϕ_t and ψ_t be two one-parameter groups, with infinitesimal generators $v(m)$ and $u(m)$, acting on a region M in $\mathbf{R}^n$. Then for small t, we have

$$\phi_t(m) = m + tv(m) + t^2\varepsilon_1(m, t), \qquad \psi_t(m) = m + tu(m) + t^2\varepsilon_2(m, t).$$

Hence

$$\begin{aligned}(\psi_t \circ \phi_t)(m) &= \psi_t(m + tv(m) + t^2\varepsilon_1(m, t)) \\ &= m + tv(m) + t^2\varepsilon_1(m, t) \\ &\quad + tu(m + tv(m) + t^2\varepsilon_1(m, t)) + t^2\varepsilon_2 \\ &= m + t(v(m) + u(m)) + t^2\varepsilon(m, t).\end{aligned} \tag{2.3.3}$$

Thus $t \to \psi_t \circ \phi_t(m)$ is a curve whose tangent vector at m is $v(m)+u(m)$. Of course, $t \to \psi_t \circ \phi_t$ is not usually a one-parameter group, but this calculation leads us to hope that, if $\{\phi_t\}$ and $\{\psi_t\}$ are subgroups of the group G of (2.3.1), then there would also be within G a one-parameter group with $v+u$ as infinitesimal generator. It is indeed so. In the case when G is the group $\mathrm{GL}(V)$ of a real vector space V (or any closed subgroup thereof), this is guaranteed by the Trotter product formula **[Howe7]**:

$$\exp(A + B) = \lim_{n\to\infty} (\exp(A/n)\exp(B/n))^n. \tag{2.3.4}$$

The correspondence between "infinitesimal composition" and addition of infinitesimal generators is very nice, but it leaves us with an enigma. Vector addition is a very faceless operation; for example, the only isomorphism invariant of vector spaces is their dimension. The simple-minded operation of vector addition cannot begin to reflect the extremely rich possibilities for group laws of Lie groups.

Thus we need to put more structure on our infinitesimal generators. A way to do this is suggested by the observation that an obvious way in which vector addition fails to capture general group laws is that it fails to be non-commutative. We could thus ask for a way to reflect the noncommutativity of a group in the infinitesimal generators of its one-parameter subgroups.

A plausible way to do this is to study the commutators of one-parameter groups. This turns out to be an excellent choice. It essentially involves refining calculation (2.3.3) to second order:

$$\begin{aligned}\psi_t \circ \phi_t(m) &= \psi_t(m + tv(m) + t^2\varepsilon_1(m, t)) \\ &= m + tv(m) + t^2\varepsilon_1(m, t) + tu(m + tv(m) + t^2\varepsilon_1(m, t)) \\ &\quad + t^2\varepsilon_2(m + tv(m) + t^2\varepsilon_1(m, t)) \\ &= m + tv(m) + tu(m) + t^2\partial_{v(m)}u(m) + t^2\varepsilon_1(m, s) \\ &\quad + t^2\varepsilon_2(m, t) + t^3\eta,\end{aligned} \tag{2.3.5}$$

where η is an appropriate smooth function and

$$\partial_{v(m)}(u)(m) = \lim_{t\to 0}\frac{u(m + tv(m)) - u(m)}{t} \tag{2.3.6}$$

is the directional derivative of u at m in the direction of $v(m)$. The term $\partial_{v(m)}(u)(m)$ is not the only second order term, but it is the first term which reflects the interaction between ψ_t and ϕ_t, and in particular is the only second order term which depends on the order of composition of ϕ_t and ψ_t. Thus, when we compute the commutator, we find

$$\psi_t \circ \phi_t \circ \psi_{-t} \circ \phi_{-t}(m) = m + t^2(\partial_{v(m)}(u)(m) - \partial_{u(m)}(v)(m)) + t^3\tilde{\eta}. \tag{2.3.7}$$

Thus, although the curve $t \to \psi_t \circ \phi_t \circ \psi_{-t} \circ \psi_{-t}$ is not necessarily smooth (it may have a cusp at $t = 0$), its geometric tangent vector at $t = 0$ is

$$\partial_{v(m)}(u)(m) - \partial_{u(m)}(v)(m). \tag{2.3.8}$$

A more rigorous result valid for pairs of matrices is the commutator formula **[Howe7]**:

$$\lim_{n\to\infty}(\exp(A/n)\exp(B/n)\exp(-A/n)\exp(-B/n))^{n^2} = \exp[A, B], \tag{2.3.9}$$

where

$$[A, B] = AB - BA \tag{2.3.10}$$

is the *commutator* of A and B.

We can see from (2.3.3) and (2.3.7), in a formal way (which was good enough for Lie), that if we have a group with the structure of a differentiable manifold acting on another manifold, then the set of infinitesimal generators of its one-parameter groups should be a vector space endowed with an antisymmetric product, given by (2.3.8) (which is now generally referred to as the *Lie bracket* of the vector fields u and v). With more work than we have done here, this can be shown rigorously. From formulas (2.3.4) and (2.3.9), we can confidently make the more modest assertion that the set of infinitesimal generators (in the sense of formula (2.2.7) and the remark following) of one-parameter groups of a closed subgroup of $\mathrm{GL}_n(\mathbf{R})$ forms a linear subspace of the space $M_n(\mathbf{R})$ of $n \times n$ matrices, and is closed under the commutator operation (2.3.10). In other words, the set of infinitesimal generators of a Lie group (of the concrete sorts we have been discussing) forms what is now called a Lie algebra. (This terminology was introduced by Hermann Weyl; the original term was "infinitesimal group.")

2.4. The incredible thing is that this bilinear product, the Lie bracket or commutator, virtually determines the group that gives rise to it. When one considers this, and then the tight control that Lie theory exercises over finite group theory (briefly described in §1.5), the tight control that finite reflection groups exercise over Lie theory (briefly described below in §§2.9, 2.10), and the manifold applications of Lie theory within mathematics and to physics (see §§3.1 and 4), it is hard to avoid a sense of awe.

To see how the commutator controls the group law, consider the following elementary calculations in $M_n(\mathbf{R})$. For matrices A, B, set

$$\begin{aligned} L_A(B) &= AB, \qquad R_A(B) = BA, \\ \mathrm{ad}_A(B) &= [A, B] = (L_A - R_A)(B). \end{aligned} \tag{2.4.1}$$

Observe that the maps

$$L: A \to L_A, \qquad R: A \to R_A \tag{2.4.2}$$

are, respectively, a homomorphism and an antihomomorphism of $M_n(\mathbf{R})$ into $\mathrm{End}(M_n(\mathbf{R}))$. In particular, if P is any polynomial in one variable, then

$$P(L_A) = L_{P(A)}, \qquad P(R_A) = R_{P(A)}. \tag{2.4.3}$$

These identities extend to convergent power series. In particular,

$$L_{\exp(tA)} R_{\exp(tB)}: C \to \exp(tA)\, C \exp(tB), \qquad C \in M_n(\mathbf{R}), \tag{2.4.4}$$

is a one-parameter group of linear transformations of $M_n(\mathbf{R})$. One easily computes that its infinitesimal generator is

$$L_A + R_B. \tag{2.4.5}$$

Taking $B = -A$ gives the famous formula

$$\exp(tA)\, C \exp(tA)^{-1} = \exp(\mathrm{ad}_A)(C). \tag{2.4.6}$$

If we follow the common practice of denoting the action of GL_n on M_n by conjugation as Ad:

$$\operatorname{Ad} g(B) = gBg^{-1}, \tag{2.4.7}$$

then we can write

$$\operatorname{Ad}\exp(tA) = \exp(t\,\mathrm{ad}_A). \tag{2.4.8}$$

Consider the nth power mapping $A \to A^n$. This is a smooth mapping from $M_n(\mathbf{R})$ to itself. Consider its derivative, which we will denote by DA^n. For each point A, DA^n is a linear map from $M_n(\mathbf{R})$ to itself, defined in the standard way (cf. **[Lang2, LoSt]**, etc.),

$$DA^n(B) = \lim_{t\to 0} \frac{(A+tB)^n - A^n}{t}. \tag{2.4.9}$$

By a computation redolent of freshman calculus, we find

$$DA^n = \sum L_{A^k} R_{A^{n-k-1}} = \sum (L_A)^k (R_A)^{n-k-1}. \tag{2.4.10}$$

Multiplying by $\operatorname{ad} A$ gives

$$(\operatorname{ad} A)(DA^n) = L_{A^n} - R_{A^n}. \tag{2.4.11}$$

Taking linear combinations over various n gives

$$\operatorname{ad} A(DP) = L_{P(A)} - R_{P(A)} \tag{2.4.12}$$

for any one-variable polynomial P. This identity extends to convergent power series. In particular,

$$\begin{aligned}(\operatorname{ad} A)(D\exp A) &= L_{\exp A} - R_{\exp A} = (L_{\exp A}(R_{\exp A})^{-1} - 1)R_{\exp A} \\ &= (\exp(\mathrm{ad}_A) - 1)R_{\exp A}.\end{aligned} \tag{2.4.13}$$

Formulas (2.4.11)–(2.4.13) are simply a convenient means to express some formal identities in power series in R_A and L_A. Consequently, we may divide (2.4.13) by $\operatorname{ad} A$ to obtain

$$D\exp A = \eta(\operatorname{ad} A)R_{\exp A}, \tag{2.4.14}$$

where

$$\eta(x) = \frac{\exp x - 1}{x} = 1 + \frac{x}{2} + \frac{x^2}{6} + \cdots + \frac{x^m}{(m+1)!} + \cdots.$$

Consider a product $\exp B \exp A$ for two matrices A, B. Since exp is analytic and invertible near 1, we know that if A, B are small enough, there is an analytic function $C(A, B)$ such that

$$\exp C(A, B) = \exp B \exp A. \tag{2.4.15}$$

Differentiate (2.4.15) with respect to B near $B = 0$. This gives

$$D\exp A(\partial_B C(A, 0)) = B \exp A.$$

Using formula (2.4.14) for $D\exp A$ gives

$$\partial_B C(A, 0) = \eta(\operatorname{ad} A)^{-1}(B). \tag{2.4.16}$$

Formula (2.4.16) has a very important consequence. The maps $L_{\exp tB}$ define a one-parameter group of linear maps of $M_n(\mathbf{R})$, with associated infinitesimal generator $v(X) = BX$. Formula (2.4.16) says, if we use coordinates around 1 (the identity matrix) obtained by pushing forward the usual linear coordinates around 0 via exp (so-called exponential coordinates or canonical coordinates), then the infinitesimal generator of $L_{\exp tb}$ has the form

$$\tilde{v}(X) = \eta(\operatorname{ad} X)^{-1}(B).^2 \tag{2.4.17}$$

But we observe this vector field is expressible solely in terms of X, B, and the commutator operation. It follows that we can express the function $C(A, B)$ of (2.4.15) as a power series in multiple commutators in A and B. The terms in this power series can be found explicitly by successive differentiation of (2.4.17). The first few terms are

$$C(A, B) = A + B + \tfrac{1}{2}[B, A] + \tfrac{1}{12}([A[A, B]] + [B, [B, A]]) + \tfrac{1}{24}[B, [A, [B, A]]] + \cdots.$$

The full formula, known as the *Baker-Campbell-Hausdorff formula* can be found in many places **[Jaco1, Serr2, HaSc]**.

From formulas (2.4.14)–(2.4.17) we can make the following somewhat technical but crucial observation:

(2.4.18) Suppose $g \subseteq M_n(\mathbf{R})$ is a Lie subalgebra, i.e., a subspace of $M_n(\mathbf{R})$ closed under the commutator operation. Then, if A, B are in g and close enough to 0, the element $C(A, B)$ of formula (2.4.15) is also in g, and can be computed strictly in terms of the commutator operation in g.

Thus, in particular, if U is a small neighborhood of 0 in g, then $\exp U$ defines a "local group." For the general Lie group, [3] this can be established by appealing to Darboux's Theorem **[Ster, Chev3, Vara]**, a general qualitative result on systems of first order P.D.E., of the same vintage as Lie's work.

2.5. Let us now stand back and see what we have found out. Let G be a Lie group. By the *Lie algebra of G* we understand the set of infinitesimal generators of one-parameter subgroups of G, endowed with a structure of vector space by means of formula (2.3.3) or (2.3.4), and with the bilinear skew-symmetric Lie bracket by means of formula (2.3.8) or (2.3.10). [4] As above, we denote the Lie algebra of G by $\operatorname{Lie}(G)$. Formulas (2.3.3), (2.3.4), (2.3.7), and (2.3.9) show that the assignment $G \to \operatorname{Lie}(G)$ is functorial in the following sense. Let $\mathbf{g}$ and $\mathbf{h}$ be Lie algebras. [4] A *homomorphism* from $\mathbf{g}$ to $\mathbf{h}$ is a linear map $\alpha\colon \mathbf{g} \to \mathbf{h}$ which takes Lie bracket to Lie bracket. Let G and

H be Lie groups, and let $\gamma: G \to H$ be a (smooth) group homomorphism. Define a mapping

$$d\gamma: \mathrm{Lie}(G) \to \mathrm{Lie}(H) \tag{2.5.1}$$

by the obvious rule: if β_t is a one-parameter subgroup of G, with infinitesimal generator x, then $d\gamma(x)$ is the infinitesimal generator of the one-parameter subgroup $\gamma(\beta_t)$. Formulas (2.3.3), (2.3.4), (2.3.7), and (2.3.9) show that $d\gamma$ is a homomorphism of Lie algebras. Clearly a composition $\gamma' \circ \gamma$ of group homomorphisms gives rise to a composition of Lie algebra homomorphisms:

$$d(\gamma' \circ \gamma) = d\gamma' \circ d\gamma. \tag{2.5.2}$$

Thus we have two classes of structures—one, Lie groups, with both geometric and algebraic aspects, and the other, Lie algebras, which are purely algebraic objects. Each of these classes has a notion of homomorphism between objects, and so forms a category. We have a correspondence between the two classes of objects, taking a Lie group G to its Lie algebra $\mathrm{Lie}(G)$. This correspondence takes homomorphisms to homomorphisms and preserves composition, so it is a functor **[Jaco2]**. It follows from results of Lie or from Ado's Theorem (see Endnote 4) that $G \to \mathrm{Lie}(G)$ is surjective—every Lie algebra is the Lie algebra of some Lie group.

Thus to complete our understanding of the correspondence $G \to \mathrm{Lie}(G)$ we need to know how many different groups correspond to the same Lie algebra (or isomorphic Lie algebras). To determine this, we use a key technical lemma **[Chev3, Serr2, Ster]**.

LEMMA 2.5.3. *Suppose G is a Lie group, with Lie algebra $\mathbf{g}$, and $\mathbf{h} \subseteq \mathbf{g}$ is a Lie subalgebra. Then there is a connected Lie group H with Lie algebra $\mathbf{h}$, and an injective homomorphism $j: H \to G$ such that $dj: \mathbf{h} \to \mathbf{g}$ is simply the inclusion map.*

The delicate aspect of this result, of course, is that $j(H)$, the subgroup generated by $\exp \mathbf{h}$, may not be closed in G. Lines of irrational slope in a torus are the familiar example. Thus, to have the correct topological structure, H must be constructed outside of G, then injected into G. The proof of Lemma 2.5.3 is an elaboration of observation (2.4.18). The argument requires some care and is somewhat tedious, but is basically straightforward.

Now consider two connected Lie groups G_1 and G_2 whose Lie algebras are isomorphic. We will abuse notation and denote them by the same letter, $\mathbf{g}$. Then the Lie algebra of the product group $G_1 \times G_2$ is just the sum $\mathbf{g} + \mathbf{g}$ of two copies of $\mathbf{g}$. The diagonal

$$\mathbf{g}_\Delta = \{(x, x) : x \in \mathbf{g}\} \tag{2.5.4}$$

is a Lie subalgebra of $\mathbf{g} + \mathbf{g}$. Let G_Δ be the group and $j: G_\Delta \to G_1 \times G_2$ be the homomorphism provided by Lemma 2.5.3 for the subalgebra $\mathbf{g}_\Delta$.

Let p_i be the projection map from $G_1 \times G_2$ onto G_i, $i = 1, 2$. The composition $p_i \circ j$ is a homomorphism from G_Δ to G_i, and the associated map $d(p_i \circ j)$ is obviously an isomorphism of Lie algebras. It follows that $p_i \circ j$ is a diffeomorphism in the neighborhood of the identity. Since it is a homomorphism, we find by translating from the identity to a general point that $p_i \circ j$ is locally a diffeomorphism at every point of G. Hence $p_i \circ j$ is a covering map (cf. **[Hu, Tits, Hoch]**, etc.).

It follows that, if G_i is simply connected (cf. **[Hu, Tits, Hoch]**, etc.), then $p_i \circ j$ must be an isomorphism. If both G_i are simply connected, then $G_1 \simeq G_\Delta \simeq G_2$. Thus up to isomorphism there is a unique simply connected group with Lie algebra $\mathbf{g}$.

On the other hand, given any connected group G with Lie algebra $\mathbf{g}$, it is routine to check that the standard construction (cf. **[Hu, Hoch]**, etc.) of the universal cover $\widetilde{G}$ of G allows one to define a group structure on $\widetilde{G}$ such that the natural projection map $\pi: \widetilde{G} \to G$ is a group homomorphism. The kernel of π must be a discrete and normal subgroup of $\widetilde{G}$; a simple argument implies that a discrete normal subgroup of a connected group is central.

The above discussion has outlined the main considerations in the proof of the following theorem, which summarizes the main foundational facts of Lie theory (Lie's Theorems 1, 2, 3 and their converses **[Gilm, Tits, Vara]**, etc.).

THEOREM 2.5.5. (a) *For each Lie algebra* $\mathbf{g}$, *there is a unique (up to canonical isomorphism) connected and simply connected Lie group* $\widetilde{G}$ *with* $\mathrm{Lie}(\widetilde{G}) = \mathbf{g}$.

(b) *Further, if* $\mathbf{g}$ *and* $\mathbf{h}$ *are Lie algebras with associated connected and simply connected groups* $\widetilde{G}$ *and* $\widetilde{H}$, *and* $\beta: \mathbf{g} \to \mathbf{h}$ *is a homomorphism of Lie algebras, then there is a unique homomorphism of groups* $\alpha: \widetilde{G} \to \widetilde{H}$ *such that*

(2.5.6)
$$\begin{array}{ccc} \mathbf{g} & \xrightarrow{\beta} & \mathbf{h} \\ {\scriptstyle \exp}\downarrow & & \downarrow{\scriptstyle \exp} \\ \widetilde{G} & \xrightarrow[\alpha]{} & \widetilde{H} \end{array}$$

commutes, i.e., $\beta = d\alpha$. *Conversely, given* α *we have seen how to construct* $\beta = d\alpha$.

(c) *If* G *is another Lie connected group with Lie algebra* $\mathbf{g}$, *then*

$$G \simeq \widetilde{G}/L \tag{2.5.7}$$

where L *is a discrete subgroup of the center of* $\widetilde{G}$; *given any such* L, *the quotient group* $\widetilde{G}/L$ *is a Lie group with Lie algebra* $\mathbf{g}$.

REMARKS. (a) Parts (a) and (b) can be more cryptically summarized as: the functor $\widetilde{G} \to \mathrm{Lie}(\widetilde{G})$ from (the category of) connected, simply connected Lie groups to (the category of) Lie algebras is an equivalence of categories.

(b) Although Ado's Theorem guarantees that any Lie algebra can be embedded in $M_n(\mathbf{R})$, it is not true that any Lie group can be embedded in $GL_n(\mathbf{R})$. For a given simply connected group $\widetilde{G}$, it may happen that only proper quotients of $\widetilde{G}$ will embed in $GL_n(\mathbf{R})$, or it may happen that only $\widetilde{G}$ itself, and no proper quotients of it, will embed in $GL_n(\mathbf{R})$. For example, $SL_m(\mathbf{C})$ is simply connected. Its center is $\mathbf{Z}/m\mathbf{Z}$. Any group covered by $SL_m(\mathbf{C})$ may be embedded in $GL_n(\mathbf{C})$ for some n. The group $\mathscr{U}$ of unipotent upper triangular real matrices is simply connected. Its center is $\mathbf{R}$. No group properly covered by $\mathscr{U}$ can be embedded in $GL_n(\mathbf{R})$. The compact orthogonal group SO_m has a fundamental group equal to $\mathbf{Z}/2\mathbf{Z}$; its universal cover is the spin group $Spin_m$, constructed by means of Clifford algebras **[Huse, Jaco2]**. The symplectic group $Sp_{2m}(\mathbf{R})$ in $2m$ variables (the isometry group of a nondegenerate, skew-symmetric form, cf. §3.2, **[Helg2, Arti]**, etc.) has fundamental group $\mathbf{Z}$. No proper cover of it can be embedded in $GL_n(\mathbf{R})$ for any n.

(c) The construction which associates to a Lie group G its Lie algebra $\mathrm{Lie}(G)$ is akin to differentiation: formulas (2.3.3) and (2.3.7) show that the vector space structure on $\mathrm{Lie}(G)$ reflects first derivatives and the Lie bracket is somehow built from second derivatives. Theorem 2.5.5 reveals the remarkable extent to which the essentially linear object $\mathrm{Lie}(G)$ determines the nonlinear object G. The faithfulness with which $\mathrm{Lie}(G)$ reflects the structure of G allows one in many situations to replace a calculation on G with a much simpler calculation on $\mathrm{Lie}(G)$. This is a key to the power of Lie theory.

(d) Since typically we think of commutativity (as opposed to noncommutativity) as contributing to simplicity, it is interesting to note that in many places in the foundations of Lie theory the presence of commutativity makes life difficult. Existence of a nontrivial center is what makes Ado's Theorem difficult (since when the center is trivial, the adjoint representation ad is faithful). Theorem 2.5.5(c) makes clear the role of the center of $\widetilde{G}$ in the nonbijectivity of the correspondence $G \to \mathrm{Lie}(G)$. The failure of Lie subgroups of a given Lie group to be closed is likewise essentially an abelian phenomenon: the standard example of a winding line on a torus captures its essence.

(e) The classical exponential map $\exp: \mathbf{R} \to \mathbf{R}^{+\times}$ is of course an isomorphism of groups. Thus in Lie theory the distinction between the additive group $\mathbf{R}$ and the multiplicative group $\mathbf{R}^{\times}$ is blurred. This blurring is essential to the theory. However, in the theory of algebraic groups, the distinction between additive groups, out of which one builds unipotent groups, and multiplicative groups, which are associated with full reducibility, is very important.

2.6. The first application of the theory summarized in Theorem 2.5.5 is to the structure of Lie groups themselves. One proves structural facts about

Lie algebras, then transfers them to Lie groups by means of Theorem 2.5.5. For example, if $\mathbf{g}$ is a Lie algebra, and $\mathbf{j} \subseteq \mathbf{g}$ is a Lie subalgebra such that $[\mathbf{g}, \mathbf{j}] \subseteq \mathbf{j}$, we say $\mathbf{j}$ is an *ideal* in $\mathbf{g}$. If G is a connected Lie group and $J \subseteq G$ is a connected normal Lie subgroup, then $\mathrm{Lie}(J)$ is an ideal in $\mathrm{Lie}(G)$; the obvious converse also holds. If $\mathbf{j} \subseteq \mathbf{g}$ is an ideal, then the quotient space $\mathbf{g}/\mathbf{j}$ inherits a natural Lie algebra structure from $\mathbf{g}$.

Here are the basic structural facts about Lie algebras. A Lie algebra $\mathbf{g}$ is *simple* if it has no ideals other than 0 and itself. The *commutator ideal* is

$$C(\mathbf{g}) = \mathbf{g}^{(2)} = [\mathbf{g}, \mathbf{g}] = \text{span of } \{[x, y] : x, y \in \mathbf{g}\}. \tag{2.6.1}$$

The *commutator series* $C^i(\mathbf{g})$ is defined by

$$C^{i+1}(\mathbf{g}) = C(C^i(\mathbf{g})). \tag{2.6.2}$$

The *descending central series* $g^{(i)}$ is defined by

$$\mathbf{g}^{(i+1)} = [\mathbf{g}, \mathbf{g}^{(i)}]. \tag{2.6.3}$$

The Lie algebra $\mathbf{g}$ is called *solvable* (in $i-1$ steps) if $C^i(\mathbf{g}) = \{0\}$ for some i; and $\mathbf{g}$ is called *nilpotent* (in $i-1$ steps) if $\mathbf{g}^{(i)} = \{0\}$ for some i. If $\mathbf{g}^{(2)} = C(\mathbf{g}) = \{0\}$, then $\mathbf{g}$ is called *commutative* or *abelian.* For a general Lie algebra $\mathbf{g}$, the *radical* of $\mathbf{g}$, $R(\mathbf{g})$, is the maximum solvable ideal in $\mathbf{g}$ (this exists); and the *nilradical* of $\mathbf{g}$, $N(\mathbf{g})$, is the maximum nilpotent ideal (this also exists). Clearly $N(\mathbf{g}) \subseteq R(\mathbf{g})$.

The reader may assume that all the terminology above is parallel to the similar group-theoretic terminology.

THEOREM 2.6.4 (cf. [**Hump, Jaco1, Serr2, Vara**], etc.). *Let* $\mathbf{g}$ *be a Lie algebra.*

(i) *We can write*

$$\mathbf{g} = R(\mathbf{g}) + \mathbf{s}_1 + \mathbf{s}_2 + \mathbf{s}_3 + \cdots + \mathbf{s}_k, \tag{2.6.5}$$

where the $\mathbf{s}_i$ *are simple nonabelian Lie subalgebras of* $\mathbf{g}$ *and the sum is direct. Then*

$$\mathbf{g}/R(\mathbf{g}) = \mathbf{s}_1 + \mathbf{s}_2 + \cdots + \mathbf{s}_k \tag{2.6.6}$$

and the $\mathbf{s}_i$ *are exactly the minimal simple ideals in* $\mathbf{g}/R(\mathbf{g})$.

(ii) $R(\mathbf{g})/N(\mathbf{g})$ *is abelian; equivalently* $R(\mathbf{g})^{(2)} \subseteq N(\mathbf{g})$.

2.7. To flesh out this structure theorem, we would like to describe $R(\mathbf{g})$ and the $\mathbf{s}_i$. It turns out that the structure of solvable Lie algebras is too flabby to permit a detailed description of all of them. However, we have a standard example of a solvable Lie algebra, namely $\mathbf{b}$, the Lie algebra of upper triangular matrices (which of course is the Lie algebra of $\mathscr{B}$, the group of invertible upper triangular matrices). (Note that the commutator ideal of $\mathbf{b}$ is $\mathbf{u}$, the Lie algebra of strictly upper triangular matrices (which is the Lie algebra of $\mathscr{U}$, the group of unipotent upper triangular matrices).) In general, we content ourselves with showing that a general solvable Lie algebra "looks like" $\mathbf{b}$ in the following sense. (For the following result we use complex scalars rather than real scalars for the same reason one uses complex scalars in discussing Jordan canonical form. Thus, let $\mathbf{b}_{\mathbf{C}}$ denote the Lie algebra of upper triangular matrices with complex entries.)

THEOREM 2.7.1 (Lie's Theorem). *Let* $\mathbf{g} \subseteq M_n(\mathbf{C})$ *be a solvable Lie subalgebra. Then* $\mathbf{g}$ *is conjugate, by an element of* $\mathrm{GL}_n(\mathbf{C})$, *to a subalgebra of* $\mathbf{b}_{\mathbf{C}}$.

REMARKS. (a) Lie's Theorem plays an important role in the representation theory of semisimple Lie algebras. See §3.5, especially Lemma 3.5.3.7.

(b) The group-theoretic version of this, known as the Lie-Kolchin Theorem **[Kolc, Serr2]**, is that a connected solvable Lie subgroup of $\mathrm{GL}_n(\mathbf{C})$ can be conjugated to be upper triangular. The generalization of this to algebraic groups is the Borel Fixed Point Theorem: a connected algebraic group acting rationally on a complete algebraic variety has a fixed point. This result plays a pivotal role in the theory of algebraic groups, especially the classification of simple algebraic groups **[Chev2]**.

2.8. In contrast to the somewhat loose situation for solvable Lie algebras, the situation for simple Lie algebras is extremely rigid, and the classification of simple complex Lie algebras, due mainly to Killing, is just 100 years old. It is an absolutely gorgeous chapter of mathematics and it continues today to inspire research. There are several excellent accounts of this currently available (cf. [**Hump, Jaco1, Serr1**], etc.). Here we will discuss its outline, in order to bring to the fore the role played by $\mathbf{sl}_2$, and to emphasize the dominant influence of the geometry of finite reflection groups.

The smallest nonabelian simple Lie algebra is $\mathbf{sl}_2$, the 2×2 matrices of trace zero (sometimes known also as the three-dimensional simple Lie algebra or TDS). It has a basis

$$(2.8.1a) \qquad e^+ = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad e^- = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad h = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

which satisfy commutation relations

$$(2.8.1b) \qquad [h, e^+] = 2e^+, \quad [h, e^-] = -2e^-, \quad [e^+, e^-] = h.$$

The algebra $\mathbf{sl}_2$ with its associated group SL_2 (the group of 2×2 matrices of determinant 1) is basic to understanding the whole family of simple Lie algebras. In fact, a careful approach to the structure theory of simple Lie algebras would first provide a detailed analysis of the linear representations of $\mathbf{sl}_2$. We will not do this (see, however, §3.5.1, especially Proposition 3.5.1.9), but let us at least remark that if we set

$$(2.8.2a) \qquad w = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \exp(\pi(e^+ - e^-)/2) \in \mathrm{SL}_2$$

then conjugation by w interchanges e^+ and e^- and reverses the sign of h:

(2.8.2b) $$\mathrm{Ad}\,w(e^+) = -e^-, \quad \mathrm{Ad}\,w(e^-) = -e^+, \quad \mathrm{Ad}\,w(h) = -h.$$

It is probably not too extravagant to say that, just as a general Lie group can be regarded as a coherent family of one-parameter groups, so a (semi)simple Lie group is a very coherent constellation of copies of SL_2. Since $\mathrm{SL}_2(\mathbf{R})$ is much less malleable than $\mathbf{R}$ itself, the corresponding list of simple Lie algebras (as opposed to all Lie algebras) is rather short and highly structured.

The basic strategy of the classification is to investigate the structure imposed on a simple Lie algebra by its action on itself by conjugation. Thus let $\mathbf{g}$ be a simple complex nonabelian Lie algebra, with Lie bracket $[\ ,\ ]$. For $x \in \mathbf{g}$, recall (cf. formula (2.4.1) or Endnote 4))

$$\mathrm{ad}\,x(y) = [x, y].$$

The first observation is this: the (generalized) eigenspace decomposition of $\mathrm{ad}\,x$ gives a grading on $\mathbf{g}$. Thus suppose y_1, y_2 are eigenvectors for $\mathrm{ad}\,x$, with eigenvalues λ_i. Using the Jacobi identity (Endnote 4) we find

$$\begin{aligned} \mathrm{ad}\,x([y_1, y_2]) &= [x, [y_1, y_2]] = [[x, y_1], y_2] + [y_1, [x, y_2]] \\ &= (\lambda_1 + \lambda_2)[y_1, y_2]. \end{aligned}$$

That is, the bracket $[y_1, y_2]$ is an eigenvector for $\mathrm{ad}\,x$, with eigenvalue $\lambda_1 + \lambda_2$. With slightly more work, one sees that if

(2.8.3) $$\mathbf{g} = \sum_\lambda \mathbf{g}(x, \lambda)$$

is the decomposition of $\mathbf{g}$ into generalized eigenspaces for $\mathrm{ad}\,x$, then

(2.8.4) $$[\mathbf{g}(x, \lambda), \mathbf{g}(x, \mu)] \subseteq \mathbf{g}(x, \lambda + \mu).$$

Thus the Jacobi identity imposes strong consistency conditions on the Lie bracket. In particular, $\mathbf{g}(x, 0)$ is a Lie subalgebra of $\mathbf{g}$, and each $\mathbf{g}(x, \lambda)$ is a module for $\mathbf{g}(x, 0)$.

To make maximum use of observation (2.8.4), we would like to find an x for which the decomposition (2.8.3) is as fine as possible. In fact, there are many such x and they are easy to find. Consider the characteristic polynomial

(2.8.5) $$\det(\mathrm{ad}(x) - \lambda I) = \sum_{i=0}^{l} \lambda^i\, \alpha_i(x), \qquad x \in g,$$

where $l = \dim \mathbf{g}$ and the $\alpha_i(x)$ are appropriate polynomials on $\mathbf{g}$. If $\alpha_i(x) = 0$ for $0 \le i < k$, but $\alpha_k(x) \ne 0$, then $\dim \mathbf{g}(x, 0) = k$ (here $\mathbf{g}(x, 0)$ is as in (2.8.3) for $\lambda = 0$). Let r be the smallest number such that the polynomial α_r on $\mathbf{g}$ is not identically zero. We call r the *rank* of $\mathbf{g}$, and we say $x \in \mathbf{g}$ is *regular* if $\alpha_r(x) \ne 0$, equivalently if $\dim \mathbf{g}(x, 0)$ is as small as possible.

Suppose $x \in \mathbf{g}$ is regular. Consider $y \in \mathbf{g}(x, 0)$. We know $\operatorname{ad} y\ (\mathbf{g}(x, \lambda)) \subseteq \mathbf{g}(x, \lambda)$ for each eigenvalue λ of $\operatorname{ad} x$. Since $\operatorname{ad} x|_{\mathbf{g}(x,\lambda)}$ is invertible for all $\lambda \neq 0$, we see that if y is sufficiently close to x, then also $\operatorname{ad} y|_{\mathbf{g}(x,\lambda)}$ is invertible. (Or, to keep the discussion algebraic, we could observe $\operatorname{ad}(x+ty)|_{\mathbf{g}(x,\lambda)}$ will be invertible for all but a finite number of scalars t.) Since $\dim \mathbf{g}(y, 0) \geq \dim \mathbf{g}(x, 0)$, if $\operatorname{ad} y|_{\mathbf{g}(x,\lambda)}$ is invertible for all $\lambda \neq 0$, we must have $\mathbf{g}(y, 0) \supseteq \mathbf{g}(x, 0)$. Since this is true for an open set of $y \in \mathbf{g}(x, 0)$, it is true for all $y \in \mathbf{g}(x, 0)$. Hence $\operatorname{ad} y|_{\mathbf{g}(x,0)}$ is nilpotent for all $y \in \mathbf{g}(x, 0)$. Hence by Engel's Theorem (cf. **[Hump, Jaco1]**, etc.) $\mathbf{g}(x, 0)$ is a nilpotent Lie algebra. It can also easily be checked to be its own normalizer in $\mathbf{g}$. Such subalgebras are called *Cartan subalgebras.* An elementary argument (using the fact that a polynomial equation $p(v) = 0$ in a complex vector space has a solution set of real codimension 2, hence the complement of the solution set must be connected), shows that, for a complex Lie algebra, all Cartan subalgebras are conjugate (by the adjoint action of the associated Lie group). Hence, although this construction appears to depend on choosing some arbitrary element of $\mathbf{g}$, in fact it is essentially canonical.

By an argument like that for finding simultaneous generalized eigenspaces for commuting operators, we find we can refine decomposition (2.8.3) to a decomposition[5]

(2.8.6) $$\mathbf{g} = \mathbf{a} + \sum \mathbf{g}_\alpha,$$

where $\mathbf{a}$ is a Cartan subalgebra of $\mathbf{g}$, and $\mathbf{g}_\alpha$ is a simultaneous generalized eigenspace for all $x \in \mathbf{a}$.

Precisely, this means that we have labeled $\mathbf{g}_\alpha$ by a linear functional $\alpha \in \mathbf{a}^*$, the dual of $\mathbf{a}$, with the property that, if $I_{\mathbf{g}_\alpha}$ denotes the identity map on $\mathbf{g}_\alpha$, then $\alpha(x) I_{\mathbf{g}_\alpha} - \operatorname{ad} x|_{\mathbf{g}_\alpha}$ is nilpotent for all $x \in \mathbf{a}$. The $\mathbf{g}_\alpha$ are called *root spaces*, and the $\alpha \in \mathbf{a}^*$, $\alpha \neq 0$, such that $\mathbf{g}_\alpha \neq \{0\}$, are called *roots.* We denote the set of roots by Σ.

To see how $\mathbf{sl}_2$ can emerge from this situation, suppose we have a pair of elements $x \in \mathbf{g}_\alpha$ and $y \in \mathbf{g}_{-\alpha}$. To keep things as simple as possible, suppose that x, y are both simultaneous eigenvectors (as opposed to generalized eigenvectors) for $\mathbf{a}$. Consider the bracket $[x, y]$. By (2.8.4) it belongs to $\mathbf{g}_0$. Consider the three (exhaustive and mutually exclusive) following possibilities:

(2.8.7)
$$\begin{aligned} &\text{(i) } \alpha([x, y]) \neq 0, \\ &\text{(ii) } \alpha([x, y]) = 0 \text{ but } [x, y] \neq 0, \\ &\text{(iii) } [x, y] = 0. \end{aligned}$$

If possibility (iii) holds, then x and y span a two-dimensional abelian subalgebra of $\mathbf{g}$. Observe that, since $\operatorname{ad} x$ and $\operatorname{ad} y$ commute and are individually nilpotent, the product $\operatorname{ad} x \operatorname{ad} y$ will also be nilpotent. If possibility (ii) holds, then $z = [x, y]$ will commute with x and y. Hence x, y, and z span a three-dimensional, two-step nilpotent Lie algebra h, commonly known as a *Heisenberg Lie algebra* (see §3.1.3). Further, we observe that since $\operatorname{ad} x$ and $\operatorname{ad} y$ are individually nilpotent, and h is nilpotent (hence solvable), the action of $\operatorname{ad} h$ on $\mathbf{g}$ consists of nilpotent operators. In particular, it follows that $\beta([x, y]) = 0$ for all roots β, not just α. Finally, suppose that (i) holds. By scaling x or y or both, we can arrange that $\alpha([x, y]) = 2$. Then comparison with formulas (2.8.1) shows that x, y, and $[x, y]$ form a standard basis for a copy of $\mathbf{sl}_2$.

Up to here, our discussion has been completely general and applies to any Lie algebra. A key point is to show that if $\mathbf{g}$ is simple and nonabelian then of the three alternatives (2.8.7), only alternative (i) is possible. The usual way to do this is via Cartan's criterion (cf. **[Hump, Jaco, Serr2, Vara]**, etc.). This involves the Killing form. This is the symmetric bilinear form on $\mathbf{g}$ defined by

$$(2.8.8) \qquad B_K(x, y) = \operatorname{trace}(\operatorname{ad} x \operatorname{ad} y), \qquad x, y \in g.$$

Cartan's criterion says that the Killing form on a simple nonabelian Lie algebra is nondegenerate. One then sees that when $\mathbf{g}$ is decomposed as in (2.8.6), the Killing form must be nondegenerate on $\mathbf{a}$, must be trivial on each $\mathbf{g}_\alpha$, and must pair $\mathbf{g}_\alpha$ and $\mathbf{g}_{-\alpha}$ nondegenerately. From these basic observations (and a thorough grasp of $\mathbf{sl}_2$) one can eliminate the occurrence of possibilities (2.8.7)(ii) and (iii). At the same time, one concludes that $-\alpha$ is a root if α is, that $\dim \mathbf{g}_\alpha = 1$ for all roots α, and (hence) that $\mathbf{a}$ is commutative and the action of $\mathbf{a}$ on $\mathbf{g}$ by ad is diagonalizable.

Thus, for each root α, one finds that the Lie subalgebra of $\mathbf{g}$ generated by $\mathbf{g}_\alpha$, and $\mathbf{g}_{-\alpha}$ is a copy of $\mathbf{sl}_2$; the interaction between these various $\mathbf{sl}_2$'s defines the structure of $\mathbf{g}$. We should remark that the proof **[Hump, Jaco1, Serr2]** of Cartan's criterion, which underlies the analysis described above, also is based on the anticipation that $\mathbf{sl}_2$ will appear inside $\mathbf{g}$ in the ways that it does. One could avoid the use of Cartan's criterion by developing more fully the consequences of the trichotomy (2.8.7). Thus at all stages in the analysis of the structure of $\mathbf{g}$, we are relying on properties of $\mathbf{sl}_2$.

2.9. To get a strong grasp on $\mathbf{g}$, we need to understand the structure of the set Σ of roots. This set turns out to have a very tight, highly symmetric structure, imposed by the $\mathbf{sl}_2$'s generated by opposing pairs $\mathbf{g}_\alpha, \mathbf{g}_{-\alpha}, \alpha \in \Sigma$, of root spaces. Let L be the subgroup of $\mathbf{a}^*$ generated by Σ; the standard name for L is the *root lattice*. The nondegeneracy of the Killing form implies that L is a discrete subgroup of $\mathbf{a}^*$, of rank equal to $\dim \mathbf{a}^*$ ($= \dim \mathbf{a} = \operatorname{rank} \mathbf{g}$). (Also, Σ spans $\mathbf{a}^*$.) Since the Killing form on $\mathbf{a}$ is nondegenerate, we can use it to identify $\mathbf{a}$ and $\mathbf{a}^*$. Then we can transfer the Killing form to $\mathbf{a}^*$, and restrict it to L. Thus L is equipped in a natural way with an inner product. Denote the dualized Killing form by B_K^*.

For $\alpha \in \Sigma$, consider the copy of $\mathbf{sl}_2$ generated by $\mathbf{g}_\alpha$ and $\mathbf{g}_{-\alpha}$ and consider the corresponding copy of SL_2 obtained by exponentiation, which acts on $\mathbf{g}$ by conjugation. Let w_α be the element in this copy of SL_2 corresponding to the element w of formula (2.8.2). Then w_α acts on $\mathbf{g}$, preserving $\mathbf{a}$, so by duality w_α acts on $\mathbf{a}^*$, preserving Σ, hence preserving L. A computation shows that

$$w_\alpha(\lambda) = \lambda - \frac{2B_K^*(\lambda, \alpha)}{B_K^*(\alpha, \alpha)}\alpha, \qquad \lambda \in L \subseteq a^*.$$

(Recall B_K^* is the dualized Killing form on a^*.) In geometrical terms, this says w_α is reflection in the hyperplane perpendicular to α. Let $W_{\mathbf{g}} = W$ be the group generated by the w_α. It is called the *Weyl group* (of the pair $(\mathbf{g}, \mathbf{a})$, or just of $\mathbf{g}$, since $\mathbf{a}$ is unique up to conjugation). Since W preserves the finite set of roots Σ, it must be a finite group. For the example $\mathbf{g} = \mathbf{sl}_n$, as described in Endnote 5, the group W is just S_n, the symmetric group on n letters.

Thus we have associated to $\mathbf{g}$ a finite group W, which is generated by reflections, and which acts on a lattice L, preserving a distinguished finite set Σ. These very elementary data determine $\mathbf{g}$.

2.10. In fact, just the group W, acting not on the lattice L but on its real span $\mathbf{a}_{\mathbf{R}}^*$, comes very close to determining $\mathbf{g}$, and the classification of finite groups of orthogonal transformations generated by reflections is very beautiful and intimately related to the classification of simple Lie algebras. Since most accounts of the classification mix together the Weyl group and the root system, we would like to make explicit here how much depends on the Weyl group alone.

The idea behind the classification of finite reflection groups is as elementary as it is elegant. Also, it is geometric to its core. Let W be a finite group acting on $\mathbf{R}^n$, and generated by reflections in hyperplanes. Let $R \subseteq W$ denote the set of reflections. For each reflection $r \in R$, let H_r be the hyperplane fixed by r. We call the H_r the *reflection hyperplanes* of W. The set $\mathbf{R}^n - \bigcup_{r\in R} H$ obtained by deleting the H_r is a finite union of open convex cones. One such cone C is called an open *Weyl chamber*; its closure $\overline{C}$ is called a closed Weyl chamber. Choose one such Weyl chamber C_0, and call it the *fundamental chamber*. The intersection of $\overline{C}_0$ with the hyperplane H_r will be some closed cone in H_r. Call H_r a *face plane* of C_0 if $H_r \cap \overline{C}_0$ has relative interior in H_r. The intersection $H_r \cap \overline{C}_0$ will be called a *face* of $\overline{C}_0$ (or of C_0). An easy argument shows that the reflections in the face planes of C_0 generate W. (More precisely, if C is any other Weyl chamber, and a line from a general point of C_0 to a general point of C passes through l hyperplanes, then C_0 can be moved to C by a product of l reflections in the faces of C_0. Hence the group generated by reflections in the face planes of C_0 acts transitively on the Weyl chambers, hence contains reflections in all hyperplanes, hence equals W. A more careful argument, proceeding by induction on word length, shows that W acts simply transitively on the Weyl chambers [**Bour**, **Hilr**].)

Thus we want to understand the relations between the reflections in the various face planes of C_0. Consider two face planes H_r and H_s of C_0. Then $H_r \cap H_s$ has codimension 2, and the group generated by r and s factors to the plane $\mathbf{R}^n/(H_r \cap H_s)$. The planar situation can easily be completely analyzed. The reflections r and s generate a dihedral group of order $2m$, $m \geq 2$, and the lines $L_r = H_r/(H_r \cap H_s)$ and $L_s = H_s/(H_r \cap H_s)$ meet at an angle π/m. See Figure 2.10.1

All the lines in the figure are the images modulo $H_r \cap H_s$ of hyperplanes $H_{r'}$, $r' \in R$, since they are transforms by the group generated by r and s of L_r and L_s. The angle between any two adjacent lines (which is also the dihedral angle between the corresponding hyperplanes) is always π/m. If H_r and H_s are both to bound a common Weyl chamber, the lines L_r and L_s must be adjacent.

In particular, the dihedral angle between H_r and H_s is always acute or a right angle. Thus, if u_r and u_s are the normal unit vectors to H_r and H_s, pointing outward from C_0, the angle between u_r and u_s is obtuse, a fact which can be expressed by saying that the dot product $u_r \cdot u_s$ is nonpositive. From this observation, an easy argument shows that the set of all u_r, for H_r a face of C_0, are independent. Thus, if we assume, without essential loss of generality, that there are no vectors fixed by all of W, the vectors u_r, for H_r a face plane of C_0, form a basis for $\mathbf{R}^n$. Thus $\overline{C}_0$ is a simplicial cone; precisely, it is the cone generated by the vectors $-u_r^*$, where u_r^* is the basis of $\mathbf{R}^n$ dual to the basis u_r, H_r a face plane of C_0. Since the dihedral angles between the faces of C_0 are acute, we call C_0 an *acute simplicial cone*. This geometry of a Weyl chamber is important in other places besides the classification of simple Lie algebras. For example, it is a key ingredient in the Langlands-Vogan classification of irreducible admissible representations [**Knap2**, **Vogal**, **Wall**] (cf. §3.6.4).

We now have the key to the classification of finite reflection groups. Since the external unit normals u_r to the faces of $\overline{C}_0$ are a basis for $\mathbf{R}^n$, the

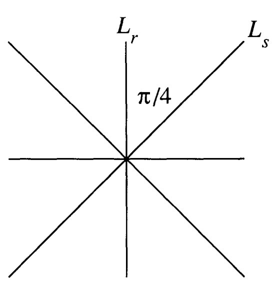


Figure 2.10.1

geometry of C_0 and hence W itself, is entirely determined by the inner products $u_r \cdot u_s$ between pairs of external normals. On the other hand, we have seen from our analysis above of the planar case that these inner products are related to the structure of W via the formula

(2.10.2) $$u_r \cdot u_s = -\cos(\pi/m_{rs}),$$

where m_{rs} is the order of the product rs (see Figure 2.10.1).

The problem is to find out what the numbers m_{rs} can be. An obvious restriction is that the Gram matrix of the u_r's, whose entries are the inner products $u_r \cdot u_s$, should be positive definite. Here yet another miracle occurs: this simple necessary condition is sufficient to completely determine all possibilities for W. Moreover, the list of possibilities is quite short, and the computations necessary to limit the list to the actual possibilities are quite easy **[Coxe, GrBe]**.

The result is usually expressed in terms of *Coxeter graphs.* For each face plane H_r of C_0 one creates a node; then the nodes for H_r and H_s are connected by $m_{rs} - 2$ lines. Alternatively, if $m_{rs} > 3$, one labels the line between node r and node s by the number m_{rs}. If two nodes are not connected, then the corresponding reflections commute with each other. If the Coxeter graph of W is disconnected, then W is a direct product of the group corresponding to the two pieces. Thus it is only necessary to record the connected Coxeter graphs. Doing so produces the list in Figure 2.10.3.

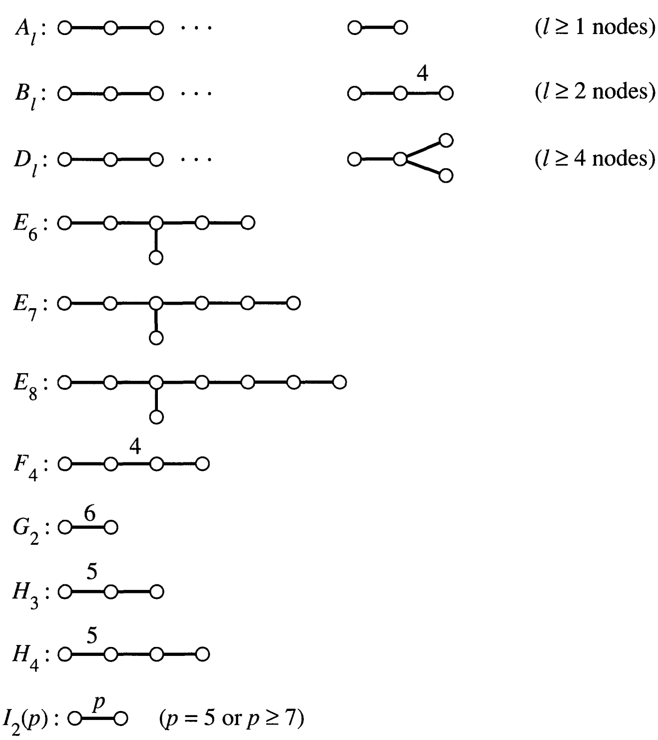


Figure 2.10.3. List of Coxeter graphs.

From our discussion of simple Lie algebras we know that their Weyl groups are contained in this list. However, not all these groups are Weyl groups, because, as we saw, the Weyl groups must leave invariant a lattice L, and not all the groups in the list above do this.[6] Additionally, for a simple Lie algebra $\mathbf{g}$, we have the data of the root system Σ (see (2.8.6) et infra). From the discussion of the connection between the root system Σ and its Weyl group W, we see that the elements of Σ are normal to the reflection hyperplanes. But since Σ is contained in the lattice L, the elements of Σ may not be unit vectors. Instead, they are characterized as the shortest vectors in L normal to the reflection hyperplanes of W.[7] It turns out (from considerations of conjugacy) that for the simple root systems only two root lengths are possible, and a change of root length can occur only between nodes which are connected by an even number of lines. To record this extra structure, one refines the Coxeter graph to what is usually called the *Dynkin diagram*, which puts an arrow across junctions with even numbers of lines, pointing in the direction of the longer roots. The resulting list (see Figure 2.10.4) contains four infinite sequences, corresponding to classical groups, and five more "exceptional" groups.

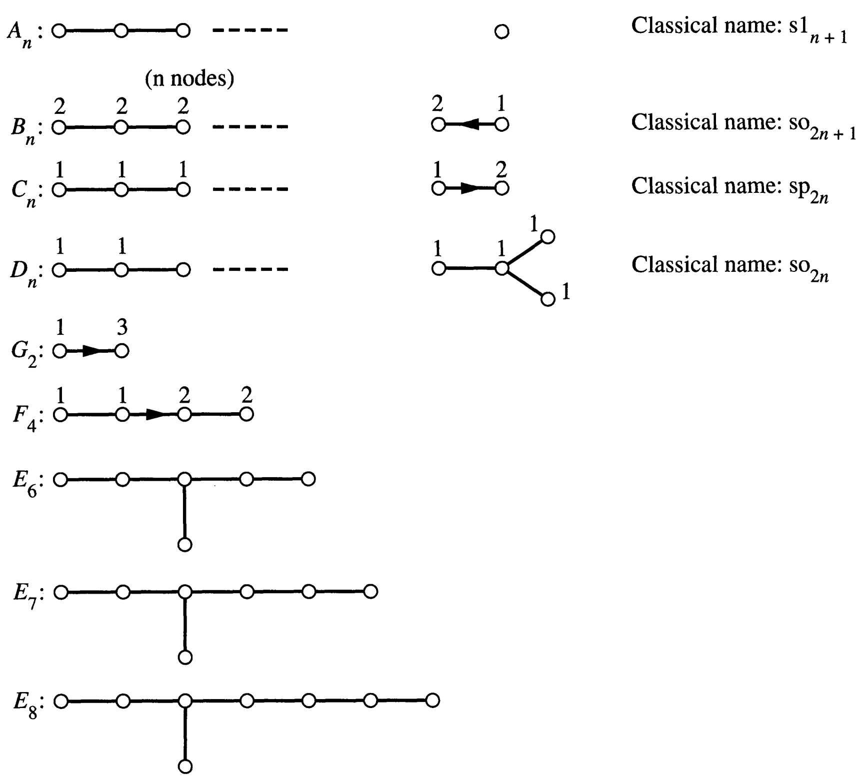


FIGURE 2.10.4. List of Dynkin diagrams.

2.11. We have travelled quite a long way in this brief tour of Lie theory. Let us sum up the major landmarks by means of a flow diagram (Figure 2.11.1).

Here rectangular blocks contain the main classes of objects involved in the theory. Ovals contain important structural information. Arrows proceeding between boxes indicate flow of information. Arrows going in opposite directions between two boxes means the objects in the boxes are each recoverable from the other.

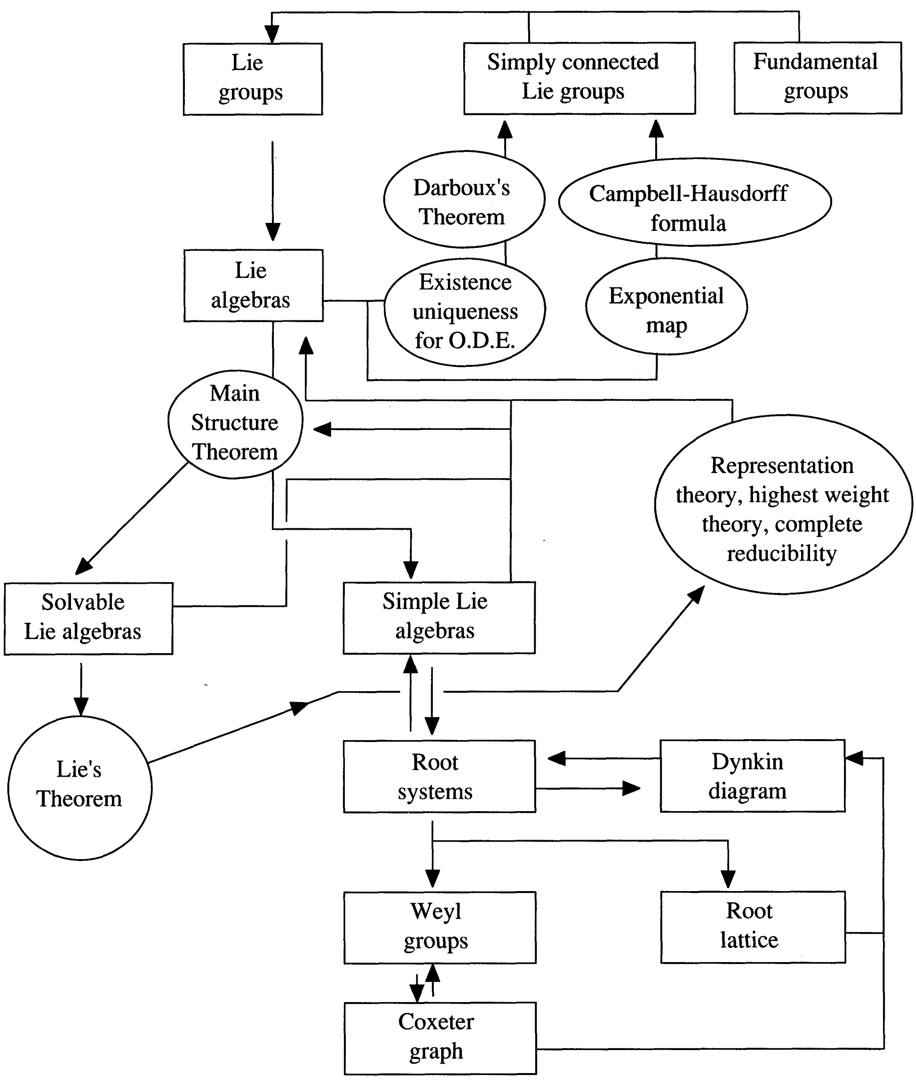


Figure 2.11.1

2.12. To emphasize that a given simple Lie algebra is recoverable from its root system (understood as a certain generating subset of a lattice L equipped with an inner product) we state the theorem of Serre [**Hump, Serr1**] describing the structure of simple $\mathbf{g}$ in terms of generators and relations based on the structure of the root systems.

The data we need are three sets of symbols X_α, Y_α, and H_α. Each triple is a standard basis for a copy of $\mathbf{sl}_2$. The triples are labeled by a set of *fundamental roots*: one chooses a fundamental Weyl chamber in $\mathbf{a}^*_{\mathbf{R}}$, and the root vectors perpendicular to the face planes of the Weyl chamber are the fundamental roots. For each fundamental root α, one chooses $X_\alpha \in \mathbf{g}_\alpha$ and $Y_\alpha \in \mathbf{g}_{-\alpha}$ such that if $H_\alpha = [X_\alpha, Y_\alpha]$, then these three elements of $\mathbf{g}$ are a standard basis of $\mathbf{sl}_2$ (see the discussion following (2.8.7)). To describe how these $\mathbf{sl}_2$'s fit together, we at least need to know the eigenvalues of the X_β under the action of H_α: set

$$\operatorname{ad} H_\alpha(X_\beta) = n_{\alpha\beta} X_\beta. \tag{2.12.1}$$

From the representation theory of $\mathbf{sl}_2$, one knows the $n_{\alpha\beta}$ are integers; and of course $n_{\alpha\alpha} = 2$. For $\alpha \neq \beta$, the integer $n_{\alpha\beta}$ is nonpositive, and is computed in terms of the geometry of the root system [**Hump, Jaco1, Serr1**]. The array of integers $\{n_{\alpha\beta}\}$ is called the *Cartan matrix* of $\mathbf{g}$ or of Σ. Serre's result says this is essentially all the data we need to specify $\mathbf{g}$.

THEOREM 2.12.2 (Serre). *Let Σ be the root system of a simple Lie algebra $\mathbf{g}$. Let $F = \{\alpha\}$ be a set of fundamental roots for Σ, and let $n_{\alpha\beta}$ be defined by equation* (2.12.1). *Let $\{X_\alpha, Y_\alpha, H_\alpha : \alpha \in F\}$ be a set of symbols. Let $\tilde{\mathbf{g}}$ be the Lie algebra generated by the X_α, Y_α, and H_a, subject to the following commutation relations*:

$$\begin{aligned} &\text{(a) } [X_a, Y_\alpha] = H_\alpha, \ [H_\alpha, H_\beta] = 0, \\ &\quad\ [X_\alpha, Y_\beta] = 0, \ \alpha \neq \beta, \\ &\text{(b) } [H_\alpha, X_\beta] = n_{\alpha\beta} X_\beta, \ [H_\alpha, Y_\beta] = -n_{\alpha\beta} Y_\beta, \\ &\text{(c) } \operatorname{ad} X_\alpha^{1-n_{\alpha\beta}}(X_\beta) = 0, \ \operatorname{ad} Y_\alpha^{1-n_{\alpha\beta}}(Y_\beta) = 0. \end{aligned} \tag{2.12.3}$$

Then $\tilde{\mathbf{g}} \simeq \mathbf{g}$.

2.13. REMARKS. (a) Descriptions, much more involved than Theorem 2.12.2 but in a similar spirit, of the Chevalley groups (over various fields) associated to simple Lie algebras have been given [**Crtr, Stei1, 2**].

(b) Serre's Theorem has assumed considerable significance in connection with Kac-Moody Lie algebras. In [**Kac2**] and [**Mood1**], Kac and Moody independently observed that one could take a "generalized Cartan matrix" of integers $n_{\alpha\beta}$ (satisfying $n_{\alpha\alpha} = 2$, $n_{\alpha\beta} \leq 0$) and define a Lie algebra $\tilde{\mathbf{g}}$ by means of relations (2.12.3). The resulting Lie algebras are infinite dimensional unless the $n_{\alpha\beta}$ came from the known list of finite-dimensional simple Lie algebras, but they share many of the important properties of finite-dimensional simple Lie algebras. In particular, each Lie algebra has an associated root system Σ (which will usually be infinite) and a Weyl group W, which is a reflection group with respect to a (possibly indefinite, even degenerate) inner product. Somewhat later, these infinite-dimensional algebras, especially the "affine" ones, whose associated inner product is positive semidefinite, were realized to be related with a range of fascinating phenomena, including power series identities (the Macdonald identities, Rogers-Ramanujan identities, etc. [**Macd2**, **Kac4**, **Kost3**, **LeMi**, **Lepo2**]); completely integrable Hamiltonian systems (Korteweg-de Vries equation, Toda lattice, etc. [**AdvM**, **DJKM1**, **2**, **GoWa1**, **2**, **Syme1**, **2**, **Kost2**]); the Fischer-Griess "Monster," the largest sporadic group [**CoNo**, **FrLM**, **Kac5**]; the representations of graphs [**DlRi**, **Ring**, **Gabr**, **Kac6**, **7**], etc.; and two-dimensional conformal field theories [**BePZ**, **Gawe**, **Witt**]. Work on these various topics is currently proceeding at a furious pace.

(c) Reflection groups, and especially root systems, figure significantly in a variety of contexts outside the classification of simple Lie algebras, some of them quite surprising. We will list a sample of these appearances.

(i) Coxeter [**Coxe**] was interested in reflection groups because of their connection with regular polytopes. It has long been understood that the symmetry groups of the platonic solids are reflection groups. The famous tesselations of the sphere associated to the regular polyhedra just show the intersection of the sphere with the Weyl chambers for the corresponding reflection group. Similarly, in higher dimensions, one can construct regular polytopes using reflection groups. Especially attractive is the four-dimensional polytope whose three-dimensional faces are 120 regular dodecahedra. Its symmetry group is H_4 in the list (2.10.3). (However, H_4 is not a Weyl group; Weyl groups are associated only to the more mundane regular solids.)

(ii) Reflection groups have an honored place in invariant theory, owing to Chevalley's theorem [**Helg**, **BeGr**, **Chev6**] complemented by Shephard and Todd [**ShTo**]:

THEOREM 2.13.1 (Chevalley, Shephard-Todd). *Let the finite group G act on the real vector space V. Let $P(V)$ be the algebra of polynomials on V, and let $P(V)^G$ be the subalgebra of polynomials invariant under the action of G. Then $P(V)^G$ is a polynomial algebra (necessarily in* $\dim V$ *variables) if and only if G is generated by reflections.*

The classical example of course is the action of the symmetric group on $\mathbf{R}^n$ by permutation of the coordinates. For this action, Lagrange's Theorem [**Jaco2**, **Lang3**, **Macd1**] says the invariant polynomials, usually called symmetric polynomials, are all expressible in terms of the "elementary symmetric polynomials"

$$\sigma_l(x) = \sum_{1 \le i_1 < i_2 < \cdots < i_l \le n} x_{i_1} x_{i_2} \cdots x_{i_l}. \qquad (2.13.2)$$

In addition to Theorem 2.13.2 there is a very beautiful description of $P(V)$ as a G-module in terms of "harmonic polynomials" **[Helg1, Chev6]**. This structure is involved in significant ways in the representation theory of semisimple groups, and the ideal theory of the universal enveloping algebras of simple Lie algebras. In particular, Theorem 2.13.2 guarantees that the center of the universal enveloping algebra of a simple Lie algebra is a polynomial ring (cf. **[Helg2, Wall2, Hump]**, etc.).

(iii) Much more recent is the application of root systems and reflection groups to problems in linear algebra defined by "representations of graphs". Let Γ be a directed graph: a collection of nodes joined by edges with a sense of direction, i.e., which proceed from one node to another node, but not backwards. (We also permit an edge to connect a node to itself.) A *representation* of the graph Γ is an assignment of a vector space V_i to each node i, and a linear transformation $T_{ij}\colon V_i \to V_j$ to each edge from i to j. There is an obvious notion of equivalence for two such representations: if $\{U_i, S_{ij}\}$ is another representation of Γ, it is equivalent to the first one if there are linear isomorphisms $J_i\colon U_i \to V_i$ such that the diagrams

$$\begin{array}{ccc} U_i & \xrightarrow{S_{ij}} & U_j \\ {\scriptstyle J_i}\downarrow & & \downarrow{\scriptstyle J_j} \\ V_i & \xrightarrow{T_{ij}} & V_j \end{array}$$

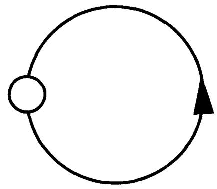
commute. There is also an obvious notion of direct sum, so the representations of Γ form an abelian category. The problem of representations of graphs is to describe (up to equivalence) the indecomposable objects in the category.

We note that several standard problems of linear algebra are formulable as graph representation problems. For example, the solution of linear equations, solved by Gaussian elimination, the essence of which is the notion of *rank* of a linear transformation, amounts to the representation problem for graph (a), and Jordan canonical form amounts to the solution of the representation problem for the one node graph (b).

Gabriel **[Gabr]** discovered that a graph Γ has only a finite number of indecomposable representations precisely when the associated undirected graph is a Dynkin diagram of type A, D, or E. This is remarkable, but the relation goes deeper: the indecomposables are naturally labeled by elements of the root system associated to the Dynkin diagram. Further, Bernstein, Gelfand, and Ponomarev **[BeGP]** showed there were functors, between representation categories of various Γ with the same Dynkin diagram as undirected graph, which imitated the action of the Weyl group. Kac **[Kac6, 7]** showed that the representations of more complicated graphs could also be analyzed in terms of root systems and Weyl groups of Kac-Moody Lie algebras.

(iv) The Dynkin diagrams (or Coxeter graphs) of types A, D, E also make

(a)



(b)

Figure 2.13.3

a fascinating appearance in algebraic geometry, in connection with the classification of isolated singularities of algebraic surfaces. To analyze such a singularity, the algebraic geometer "blows it up" until he achieves a non-singular surface **[Lauf, Hirzl, Koda]**. In the process, the singular point is replaced by a system of curves. In the case of isolated singularities, the curves are all just projective lines (Riemann spheres—we use "curve" in the sense of a complex one-dimensional (hence real two-dimensional) manifold). To describe the resulting array of curves, one constructs a graph by creating a node for each projective line, and connecting two nodes by the negative of the intersection number of the two associated curves. (A p-manifold and a q-manifold inside a $(p+q)$-manifold can be expected "generically" to intersect transversally in a finite number of points. To each point of intersection, one can associate a ± 1 according as the orientation provided by local coordinates on the submanifolds agrees or disagrees with the orientation of the ambient manifold. The sum of these ± 1's, over all points of intersection, is the intersection number.) It turns out that in the case of simple isolated singularitites, the resulting graph is always a Coxeter graph of type A_n, D_n, E_6, E_7, or E_8 **[Arno3, Looi]**.

This remarkable result has been analyzed in two different ways. One is in terms of finite subgroups of SU_2. These have been understood since Klein **[Klei, GrBe, Miln]** and are of course themselves closely related to three-dimensional reflection subgroups. For any $g \neq 1$ in SU_2, the only point of $\mathbf{C}^2$ fixed by g is the origin. Hence, for any finite $G \subseteq \mathrm{SU}_2$ the image of the origin in the quotient space $\mathbf{C}^2/G$ is an isolated singularity. All the isolated surface singularities arise in this way. Further, it is possible to recapture the Coxeter graph of the singularity directly from the representation theory of G. Let $\widehat{G}$ be the unitary dual of G—the set of its (equivalence classes of) irreducible (unitary) representations. Let ρ_0 be the given representation of G on $\mathbf{C}^2$. Define a graph whose nodes are the elements of $\widehat{G}$, and such that σ_1 and σ_2 are connected if and only if σ_2 is a component of $\sigma_1 \otimes \rho_0$.

(This is a symmetric relation.) The resulting graph is the graph associated geometrically to the singularity $\mathbf{C}^2/G$ **[Lamo, Looi]**.

There is also a direct connection between the simple Lie algebra $\mathbf{g}$ with graph of type A_n, D_n, or E_n, and the singularity with the same graph. We call an element x of $\mathbf{g}$ *nilpotent* if $\operatorname{ad} x$ is nilpotent. The set of nilpotent elements forms an algebraic subvariety η of $\mathbf{g}$, of codimension equal to the rank of $\mathbf{g}$. Let G be the Lie group associated to $\mathbf{g}$, and let Ad be the action of G on $\mathbf{g}$ by conjugation. Then η may be characterized as the set of zeros of the $\operatorname{Ad} G$-invariant polynomials which vanish at the origin. Further, η consists of only finitely many $\operatorname{Ad} G$-orbits. (For $\mathbf{g} = \mathbf{gl}_n$, these are described by Jordan canonical form.) There is one G-orbit which is open-and-dense, consisting of the so-called *regular nilpotent* elements. The complement in η of the regular nilpotent elements is the singular set of η, and has codimension 2. Denote it by η_1. If one takes a two-dimensional slice in η, at a typical point of η_1 and transverse to η_1, this two-dimensional variety will have an isolated singularity, of the type corresponding to the Coxeter graph of $\mathbf{g}$ **[Brie, Slod]**.

(v) Finally, to emphasize how innocently, and from what seemingly meager contexts, the root system of simple Lie groups can arise, consider the question of integral quadratic forms. Let V be a real vector space, $L \subseteq V$ a lattice, and $B(,)$ a (positive-definite) inner product on V, such that $B(l, l')$ is an integer if $l, l' \in L$. The question is to describe the isometry classes of such forms, modulo automorphisms of L.

Suppose $l_0 \in L$ has B-norm equal to $1 : B(l_0, l_0) = 1$. Then for any $l \in L$ the difference $l - B(l, l_0)l_0$ is orthogonal to l_0. Hence if U_0 is the line through l_0, and $U_0^{\perp}$ is the hyperplane orthogonal to U_0, then

$$L = (L \cap U_0) + (L \cap U_0^{\perp})$$

(orthogonal direct sum).

Hence for purposes of our classification problem, we may as well assume there are no vectors in L of length 1. Consider next the possibility of vectors l such that $B(l, l) = 2$. Let $L_1 \subseteq L$ be the lattice spanned by such vectors. Then L_1 decomposes into an orthogonal direct sum of lattices, each one of which is naturally isometric to the root lattice (with appropriately scaled Killing form) of one of the simple Lie algebras, of type A, D, or E; and the vectors l with $B(l, l) = 2$ form the root system of the appropriate type.

The root lattice of E_8 is particularly significant in this context. Given V, L, B as above, define

$$L^* = \{v \in V : B(v, l) \in \mathbf{Z} \text{ for all } l \in L\}. \tag{2.13.4}$$

Then L^* is a lattice, and $L^{**} = L$. By our assumption on L, we have $L \subseteq L^*$. The quotient group L^*/L is clearly an invariant of the isometry class of L. Of particular interest are the *self-dual* or *unimodular* lattices, for which $L = L^*$. Of course $\mathbf{Z}^n$, with its usual inner product, is self-dual;

but a more interesting problem is to find an *even* unimodular lattice, i.e., one for which $B(l, l)$ is even for all $l \in L$. It turns out there are none in dimensions less than 8, and that the root lattice of E_8 is the unique example in dimension 8 **[FrLM, Sloa, CoSh]**.

To reinforce the opinion that the facts just recited are not merely curiosities, but are worthy of contemplation, we recall that the Leech lattice, which is the unique even unimodular lattice in 24 dimensions such that $B(l, l) \geq 4$ for all $l \neq 0$, is deeply involved with the sporadic simple groups, especially the Conway groups and the Monster **[CoNo, FrLM, Thom]**.

(d) The classification of simple Lie algebras over fields of positive characteristic is much more delicate than in characteristic zero, because of the failure of Lie's Theorem (Theorem 2.7.1) and related problems. Although, the last word has not been said on this, nearly the last is contained in **[StWi]**, which shows that the types $A - G$, plus a family of other algebras, analogous to Lie algebras in characteristic zero which are infinite dimensional, constitute all simple Lie algebras over an algebraically closed field of characteristic $p \geq 7$.

Endnotes. 1. In other words, $t^+(m) = +\infty$ and $t^-(m) = -\infty$ for all m. Note that equation (2.1.8) implies that $t^{\pm}(\gamma_s(m)) = t^{\pm}(m) - s$. Hence if there exists $\varepsilon > 0$ such that $t^-(m) \leq -\varepsilon$ and $t^+(m) \geq \varepsilon$ for all points m, then $t^{\pm}(m) = \pm\infty$ for all m. That is, if we can solve (2.1.6) in a uniform interval for all initial conditions, we can solve it for all time.

2. The function

$$\eta(x)^{-1} = \frac{x}{e^x - 1} = \sum_{n=0}^{\infty} B_n\, x^{2n} - \frac{x}{2},$$

where the B_n are the Bernoulli numbers (cf. **[Hirz, Lang5]**, etc.).

3. Some readers may be bothered by the fact that we have not given a formal definition of Lie group. We present one here for them. A Lie group is a smooth manifold G endowed with a group structure such that the maps

$$\begin{aligned} G \times G &\to G, & G &\to G, \\ (x, y) &\to xy, & x &\to x^{-1} \end{aligned}$$

of multiplication and inversion (or, equivalently, the single map $(x, y) \to xy^{-1}$) are smooth. Clearly, by letting G act on itself by left translations, we can realize G as a group of diffeomorphisms of a smooth manifold.

4. The question presents itself: will any skew-symmetric product on a vector space define a Lie algebra, in the sense that it arises as the set of infinitesimal generators of a Lie group? The answer is negative. There is an additional identity that needs to be satisfied, the *Jacobi identity*:

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0.$$

This is easy to verify for either of our concrete Lie algebras (of vector fields or matrices). It is an infinitesimal analog of the associative law for group multiplication.

Lie showed that given a vector space with a skew-symmetric product satisfying the Jacobi identity he could construct a group (or what is now called a "local group") with that Lie algebra. Later Ado **[Hump, Jaco, Vara]**, showed a vector space with a skew-symmetric product satisfying the Jacobi identity could be realized as a Lie algebra of matrices, i.e., a subspace of $M_n(\mathbf{R})$, closed under the commutator operation. Thus we see that a satisfactory definition of abstract Lie algebra, without reference to a group, is: a vector space endowed with a skew-symmetric bilinear form satisfying the Jacobi identity.

The efficiency with which the Jacobi identity captures the essence of Lie algebra structure is shown by the following two observations. First, given a Lie algebra $\mathbf{g}$ with bracket operation $[\,,\,]$, define

$$\operatorname{ad} x : \mathbf{g} \to \mathbf{g},$$
$$\operatorname{ad} x(y) = [x, y], \qquad x, y \in \mathbf{g}.$$

Then $\operatorname{ad}: \mathbf{g} \to \operatorname{End}(\mathbf{g})$ is a linear map. The Jacobi identity says

$$[\operatorname{ad} x, \operatorname{ad} y] = \operatorname{ad}[x, y]$$

(where the $[\,,\,]$ on the left-hand side is the commutator of operators (2.3.10)). That is, the map ad preserves Lie bracket, and so is a representation of $\mathbf{g}$ on itself. (From formula (2.4.8), we see ad is the infinitesimal version of conjugation.) In particular, if $\mathbf{g}$ has no center (i.e., no nonzero elements x such that $[x, y] = 0$ for all $y \in \mathbf{g}$), then ad gives an isomorphism of $\mathbf{g}$ with a Lie algebra of matrices (Ado's Theorem for such $\mathbf{g}$). In particular, if $\mathbf{g}$ is a nonabelian simple Lie algebra, then ad provides a faithful matrix representation of $\mathbf{g}$.

Second, the Jacobi identity also says

$$\operatorname{ad} x([y, z]) = [\operatorname{ad} x(y), z] + [y, \operatorname{ad} x(z)].$$

This says that $\operatorname{ad} x$ is a derivation **[Lang3]** of $\mathbf{g}$. It follows from purely formal properties of exp that $\exp(\operatorname{ad} x)$ must be an automorphism. Thus a Lie algebra structure always comes with a group of automorphisms, provided by conjugation by the associated group. This implies that the more complicated the Lie algebra structure, the more symmetrical it must be. If this observation is too vague to produce the very clean list of simple Lie algebras, at least it is consistent with the existence of such a list.

5. The standard example is $\mathrm{gl}_n = M_n(\mathbf{R})$. One can take for $\mathbf{a}$ the diagonal matrices. Then the $\mathbf{g}_\alpha$'s are the lines generated by the matrix units E_{ij} (cf. formula (1.2.1)). If we use the usual coordinates a_i on $\mathbf{a}$, then the roots α are $a_i - a_j$.

6. This issue, like so many others, is settled simply by considering the pairwise relations between generators. It is easy to see that a reflection group in the plane generated by r and s can only preserve a lattice if $m_{rs} =$

2, 3, 4, or 6. Comparison of lists (2.10.3) and (2.10.4) reveals it is exactly the Coxeter graphs which have an m_{rs} other than 2, 3, 4, or 6 which do not survive to become Dynkin diagrams.

7. This is for "reduced" root systems, which is what is encountered in classifying simple complex Lie algebras. For real Lie algebras, nonreduced root systems, e.g., BC_n, can also occur **[Helg2, Serr1]**.

3. Representation theory. Research into representations (actions on vector spaces via linear transformations) of Lie groups, motivated on one hand by physics **[FlSz, Mack1, ITGT1–17, Barg3]** and on the other by the theory of automorphic forms **[GGPS, JaLa, Weil1, BoCa]** with deep roots in classical analysis and with strong ties to differential equations, and of course also propelled by its internal dynamics, has been a major part of the mathematical enterprise since roughly World War II. Considering the diversity of motivations, goals, people, and methods involved, the subject displays a remarkable amount of unity. A major source of the unity is the philosophy of the *orbit method* (also known by the more fashionable term *geometric quantization* **[Blat, Kiri, Kost1, Sour]**). Although we can only sample from the wide range of results that have been established, the overall coherence provided by the viewpoint of the orbit method allows us to convey much more of the subject than would otherwise be possible. An interesting technical point, however, is that the orbit method is almost exclusively a method of *interpretation*, a way of organizing results into a coherent (and often very beautiful) pattern. It provides little in the way of technical tools for proofs or computations. Thus, for example, several of the major results of Harish-Chandra on representations of semisimple groups have found elegant interpretations in terms of the orbit method **[Ross1, 2, DuVe, DuHV]**. However, these interpretations have provided no short-cuts to Harish-Chandra's proofs of these results.

A proper discussion of representation theory requires an aggravatingly long technical preparation. We are going to try to ignore that here. For the convenience of the reader, basic definitions and constructions have been summarized in Appendix 1. The discussion below refers to Appendix 1 as necessary. The reader who finds these references too distracting may wish to acquaint himself, at least in a rough way, with Appendix 1 before reading the main body of this section.

3.1. *An example: the quantum harmonic oscillator.* To illustrate the potential uses of representation theory, and its attraction, I can produce no better example than the spectral analysis of the quantum mechanical harmonic oscillator. This is elementary almost to the point of simple-mindedness, yet it contains the seeds of extremely varied developments that form subjects of active current research. In particular, it is basic for the orbit method to be discussed later. Also, it exhibits the extreme elegance of the best Lie algebraic computations.