SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Semisimple Lie algebras and the Cartan decomposition

av

Martin Nilsson

2018 - No K20

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

Semisimple Lie algebras and the Cartan decomposition

Martin Nilsson

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Gregory Arone

2018

Semisimple Lie algebras and the Cartan decomposition

Martin Nilsson

June 6, 2018

Abstract

Consider a set of matrices that is closed under both linear combinations and the "commutator" $AB - BA$ of any pair of matrices A, B of the set. This is what is known as a linear Lie algebra; these generalize to abstract Lie algebras, which possess a commutator-like operation but need not consist of matrices. We begin with a brief discussion on how Lie algebras arise, followed by an investigation of some basic properties of Lie algebras and what can be said in the general case. We then turn to semisimple Lie algebras—those that can be built up from "simple" ones—and study in depth their representations, or ways to inject them into linear Lie algebras in a structure-preserving fashion. After deriving a sufficient breadth of results, we then proceed with exploiting a certain representation and its properties in order to deconstruct any given semisimple algebra into its so-called Cartan decomposition. Finally, we show how any such decomposition can be understood in terms of its "root system", an associated geometric object embedded in some Euclidean space.

Contents

Acknowledgements

I would like to thank my supervisor Gregory Arone for introducing me to the subject and recommending literature, for participating in helpful discussions, and for helping me proof-read the thesis. I would also like to thank Dan Petersen for his additional proof-reading and help.

1 Introduction: Lie Groups

This introductory chapter is meant to guide the reader through the construction of Lie algebras as "linearizations" of Lie groups, which is how they arise in practice. Here we assume familiarity with concepts from differential geometry, though one can safely skip to to the next section without losing any necessary theory as long as one is willing to accept Lie algebras at their "face value".
We start with a definition:

Definition 1.1. A **Lie group** is a smooth manifold equipped with a group structure in such a way that $(g, h) \mapsto gh$ and $g \mapsto g^{-1}$ define smooth maps $G \times G \to G$ and $G \to G$, respectively.

One can show that $\mathrm{GL}(V)$, the group of all invertible linear operators over V, is a Lie group for any finite-dimensional real vector space V (see Section 7.1 of [4]). Along with its *Lie subgroups*—that is, the subgroups that also inherits a manifold structure from its "parent"—$\mathrm{GL}(V)$ provides many canonical examples of Lie groups. We will here be content with pointing out only one of these (as Lie groups will not be our main topic of study); this subgroup is $\mathrm{SL}(V)$, the set of all linear operators on V with determinant equal to 1.

Definition 1.2.

- A map $\rho : G \to H$ between Lie groups is said to be a **homomorphism** (of Lie groups) if it is smooth and a group homomorphism.
- A map ρ as above is an **isomorphism** (of Lie groups) if it is bijective with smooth inverse.
- An isomorphism $\rho : G \to G$ is called an **automorphism of** G, and we write $\rho \in \mathrm{Aut}\ G$.

Let e denote the group identity of G. We begin with considering for any given $g \in G$ the conjugation map $\Psi_g : G \to G$, i.e. $\Psi_g(h) = ghg^{-1}$ for all $h \in G$. It is a group isomorphism from G to itself with inverse $\Psi_{g^{-1}}$. Moreover, it is a smooth map from G to itself since it is the the group inversion map followed by two applications of the group multiplication map, and these are smooth maps by assumption. Then $\Psi_{g^{-1}}$ is smooth, too, meaning Ψ_g is an automorphism of G. As g varies in G, we obtain a map

$$\Psi : G \to \mathrm{Aut}\ G,$$

and since $\Psi_{gh} = \Psi_g \circ \Psi_h$, this map is in fact a group homomorphism. Again fix $g \in G$. Since Ψ_g is smooth, we may consider its differential at the identity:

$$(d\Psi_g)_e : T_eG \to T_{\Psi_g(e)}G.$$

Let us call this differential the *adjoint representation of g in G* and write $\mathrm{ad}_G\ g$ as a shorthand. Since $\Psi_g(e) = geg^{-1} = gg^{-1} = e$,

$$\mathrm{ad}_G\ g : T_eG \to T_eG,$$

and because tangent spaces are real vector spaces and differentials are linear transformations between them, we have $\mathrm{ad}_G\ g \in \mathrm{End}\ T_eG$ for all $g \in G$, where $\mathrm{End}\ T_eG$ denotes the set of all linear operators of T_eG. We can be more specific: The differential of a smooth map with smooth inverse (which we know Ψ_g to have) is always invertible, so $\mathrm{ad}_G\ g$ is invertible. If we write $\mathrm{GL}(T_eG)$ for the group of all invertible linear operators on T_eG, we therefore have a map

$$\mathrm{ad}_G : G \to \mathrm{GL}(T_eG).$$

Let $g, h \in G$ be arbitrary. By the chain rule for differentials,

$$\mathrm{ad}_G\ gh = (d\Psi_{gh})_e = (d(\Psi_g \circ \Psi_h))_e = (d\Psi_g)_e \circ (d\Psi_h)_e = \mathrm{ad}_G\ g \circ \mathrm{ad}_G\ h,$$

and it follows that ad_G is a group homomorphism. Since both its domain and codomain are Lie groups (as for the latter, recall the discussion preceding Definition 1.2) we naturally wonder if not ad_G is also a homomorphism of Lie groups. This is in fact true, though we will not show it here. As it allows for the identification of G with a Lie subgroup of a linear group, i.e. it represents G as a group of linear operators (though perhaps with loss of information if ad_G is not injective), we call ad_G the *adjoint representation of G*.
Now, let us write $L = T_eG$. As ad_G is a smooth map, it is differentiable, and its differential at e should send elements of L into the tangent space of $\mathrm{GL}(L)$ at $\mathrm{ad}_G\ e = (d\Psi_e)_e = (d\ \mathrm{Id}_G)_e = \mathrm{Id}_L$. (To verify this last claim, let $X \in L$. Then $(d\ \mathrm{Id}_G)_e(X)(f) = X(f \circ \mathrm{Id}_G) = X(f)$ for all $f \in C^\infty(G)$, or $(d\ \mathrm{Id}_G)_e(X) = X$.) To understand this tangent space better, we make use of the fact that $\mathrm{GL}(L)$ is a smooth embedded submanifold of $\mathrm{End}\ L$, where the latter denotes the real vector space of linear operators on L. Due to this inclusion, Proposition C.3 of [5] tells us that

Remark. The tangent space to $\mathrm{GL}(L)$ at Id_L can be regarded as the set of all $\mathcal{A} \in \mathrm{End}\ L$ such that there exists a smooth curve γ in $\mathrm{GL}(L)$ with $\gamma(0) = \mathrm{Id}_L$ and $d\gamma/dt|_{t=0} = \mathcal{A}$.

and so if we disregard exactly what maps to what, we see that what we really have is a mapping

$$L \to \mathrm{End}\ L.$$

The notation is at this point starting to get cumbersome, so let us just write ad for this map. If one has a background in computer science, then it is clear that we may define a function $[-,-] : L \times L \to L$ by "uncurrying" ad:

$$[X, Y] = \mathrm{ad}(X)(Y), \quad (X, Y \in L).$$

In words, we send elements $X \in L$ to functions $\mathrm{ad}(X) \in \mathrm{End}\ L$, and then evaluate those functions at elements $Y \in L$. Now, ad is linear (it is just a differential whose image we embedded in a new codomain), meaning $[-,-]$ is linear in its first argument. But each $[X,-]$ is a linear operator on L, so we see that $[-,-]$ *is bilinear.*

So far, our discussion has taken us from our initial Lie group G to the real vector space $L = T_eG$, which we now see to be equipped with a bilinear operation $[-,-]$. By giving more attention to details, one can derive additional properties of $[-,-]$, such as it being skew-symmetric and satisfying a certain technical identity called the *Jacobi identity.* We will will be content with stopping here since this gives most of the motivation we need. For the reader, the takeaway is this: *What we have derived here is exactly a Lie algebra.*

As intuition for why the Lie algebra of a Lie group could be of value, suppose G is connected; it is a fact (see Section 8.1 of [4]) that G is generated as group by any neighbourhood of e and that any homomorphism $\rho : G \to H$ is uniquely determined by the differential $d\rho_e : T_eG \to T_eH$. This is perhaps not so surprising, as the tangent space T_eG is in some sense the "best linear approximation" of the Lie group at its identity (at least if we apply our intuition on submanifolds of $\mathbb{R}^n$), and it is natural that the additional rigorosity provided by the group structure could force this approximate relationship into a more exact one.

For our final remark of this section, we state (without proof) that

$$d\rho_e([X,Y]) = [d\rho_e(X), d\rho_e(Y)], \quad (X, Y \in L),$$

where we in the right-hand side mean the operation $[-,-]$ as defined on T_eH. This identity motivates the definition of a homomorphism between Lie algebras as a linear transformation that preserves the bilinear operation.

2 Lie algebras

Throughout the remaining text, all results and their accompanying proofs are, unless otherwise stated, based partly or completely on those of [7].

2.1 Fundamental definitions and results

Let V be a vector space over some field F, and let $\mathrm{End}\ V$ denote the set of all linear operators of V. Of course, "End" stands for endomorphism, which here means a linear mapping from V to itself, or just a linear operator of V. The set $\mathrm{End}\ V$ is itself a vector space over F under addition of operators and multiplication of operators with scalars. Moreover, $\mathrm{End}\ V$ is (in the terminology of [7]) an *F-algebra*, by which is meant a vector space over F equipped with a bilinear operation. (The reader might be more familiar with referring to such a structure as an algebra over a field). In $\mathrm{End}\ V$, the bilinear operation is given by composition of operators and is as usual denoted by juxtaposition. Note that though composition is associative, we do not in general require the bilinear operation to be associative in our definition of an F-algebra.

Now introduce a new binary operation $[-,-]$ on End V, called the *bracket*, by letting $[\mathcal{A},\mathcal{B}] = \mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A}$ for all $\mathcal{A},\mathcal{B} \in \text{End } V$. We usually suppress the comma and write $[\mathcal{A}\mathcal{B}]$ if there is no ambiguity. Clearly, $[\mathcal{A}\mathcal{B}] = 0$ if $\mathcal{A}$, $\mathcal{B}$ commute, so the bracket can be said to be a measure of the "commutativity" of pairs of operators. Equipping End V with this operation gives rise to an F-algebra, but a different one from when the operation is composition. For this to be the case, we have to verify that the bracket is bilinear. Starting with the first argument,

$$\begin{aligned}[a\mathcal{A} + b\mathcal{B}, \mathcal{C}] &= (a\mathcal{A} + b\mathcal{B})\mathcal{C} - \mathcal{C}(a\mathcal{A} + b\mathcal{B}) \\ &= a(\mathcal{A}\mathcal{C} - \mathcal{C}\mathcal{A}) + b(\mathcal{B}\mathcal{C} - \mathcal{C}\mathcal{B}) = a[\mathcal{A}\mathcal{C}] + [\mathcal{B}\mathcal{C}].\end{aligned}$$

Here as in the rest of the text we write $\mathcal{A},\mathcal{B},\mathcal{C},\ldots$ for operators in End V and $a,b,c,\ldots$ for scalars in F. Linearity in the other argument then follows from $[\mathcal{A}\mathcal{B}] = \mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A} = -(\mathcal{B}\mathcal{A} - \mathcal{A}\mathcal{B}) = -[\mathcal{B}\mathcal{A}]$, or in words, the anticommutative property of the bracket.

Definition 2.1. A **linear Lie algebra** is an F-subalgebra of End V, i.e. a subspace of End V closed under the bracket.

We say that a linear Lie algebra L, contained in End V, is *finite dimensional* if V is. When we view End V as a linear Lie algebra in its own right, we denote it $\mathfrak{gl}(V)$, and we use lowercase letters $x, y, z, \ldots$ to denote its elements. The only difference is that we explicitly acknowledge the additional structure provided by the bracket.
Linear Lie algebras satisfy many important and useful results, but what is remarkable is that many of these hold for a larger class of vector spaces modeled on linear Lie algebras. These are called *abstract Lie algebras*, or just Lie algebras, and their definition is given next.

Definition 2.2. Let L be an F-algebra, with bilinear operation $[-,-]$, which we call the bracket of L. We say that L is a **Lie algebra** if, for all $x, y, z \in L$,

(L1) $[xx] = 0$;

(L2) $[x[yz]] + [y[zx]] + [z[xy]] = 0$ (the *Jacobi identity*).

Both axioms are generalizations of properties satisfied by the bracket of a linear Lie algebra: (L1) is trivial in the linear case since $[xx] = xx - xx = 0$, but (L2) is not as obvious. For completeness sake we verify it below, which additionally shows that linear Lie algebras are (as expected) special examples of Lie algebras. Letting L be a linear Lie algebra, we see that for all $x, y, z \in L$,

$$\begin{aligned}&[x[yz]] + [y[zx]] + [z[xy]] = \\ &[x, yz - zy] + [y, zx - xz] + [z, xy - yx] = \\ &[x, yz] - [x, zy] + [y, zx] - [y, xz] + [z, xy] - [z, yx] = \\ &xyz - yzx - xzy + zyx + yzx - zxy - yxz + xzy + zxy - xyz - zyx + yxz = \\ &0 - 0 - 0 + 0 - 0 - 0 = 0.\end{aligned}$$

In addition, anticommutativity of the bracket of abstract Lie algebras follows from bilinearity and (L1), since these together imply

$$0 = [x+y, x+y] = [xx] + [xy] + [yx] + [yy] = [xy] + [yx], \quad (x, y \in L).$$

Rearranging this equation, we obtain $[xy] = -[yx]$.
We say that L is finite dimensional if it is finite dimensional as vector space. If L moreover happens to be a linear Lie algebra, then we (as before) impose that its underlying vector space V be finite dimensional.
As is typical of many other algebraic theories, we have notions of substructures and structure-preserving transformations between Lie algebras.

Definition 2.3. A **subalgebra** K of a Lie algebra L is an F-subalgebra of L, i.e. $[xy] \in K$ for all $x, y \in K$.

Definition 2.4. Let L,M be Lie algebras. A **homomorphism** (of Lie algebras) is an F-algebra homomorphism $\phi : L \to M$, i.e. a linear transformation such that $\phi([xy]) = [\phi(x)\phi(y)]$ for all $x, y \in L$.

Definition 2.5. A bijective homomorphism $\phi : L \to M$ is called an **isomorphism**. When an isomorphism exists between L and M, we call L and M **isomorphic**, and write $L \cong M$.

To say that L is a linear Lie algebra is hence equivalent to saying that L is a subalgebra of $\mathfrak{gl}(V)$ for some vector space V.

Remark. To shorten definitions and proofs, we will adopt a shorthand where we allow expressions to contain one or more sets where elements would usually stand, e.g. $[xL]$. Such expressions are to be understood as denoting the set of all expressions obtained when the participating set(s) are replaced with one of its elements, assuming that the resulting set is well-defined. Our previous example would therefore define the set $[xL] = \{[xy] \mid y \in L\}$, and, in a similar spirit $[LM] = \{[xy] \mid x \in L, y \in M\}$. There is one exception to this rule, however, and that is when we write $[IJ]$ for two ideals I, J of the same algebra. We explain next what we mean by an ideal, and after that, our convention for $[IJ]$.

We use this shorthand in our next few definitions. The symbol $\subset$ will be reserved for inclusions of sets and does not necessarily imply proper inclusion.

Definition 2.6. An **ideal** I of a Lie algebra L is a subalgebra of L for which $[xI] \subset I$ for all $x \in L$. That is, $x \in L$, $y \in I$ imply $[xy] \in I$.

Due to anticommutativity, it would not have mattered had we instead chosen $[Ix] \subset I$ to be the defining property for ideals. We will often make use of this fact when showing that a subalgebra under consideration is an ideal.
There are always at least two ideals in L—the zero subalgebra $\{0\}$, denoted 0, and L itself. These are the *trivial* ideals, and may of course coincide, which happens when $L = 0$. As in ring theory, if I, J are ideals, then $I \cap J$ and $I + J$ are, too, which not difficult to verify. For a third way to construct new ideals

from old, consider the set of all finite linear combinations of elements $[xy]$, where $x \in I$, $y \in J$. We denote this set $[IJ]$ (or $[I, J]$). In set-builder notation,

$$[IJ] = \{\sum_{i=1}^{n} a_i [x_i y_i] \mid n \in \mathbb{Z}^+; a_i \in F, x_i \in I, y_i \in J; i = 1, 2 \ldots, n\}.$$

Clearly, $[IJ] \subset I \cap J$, and then its definition implies that it is an ideal of L. A special case is $[LL]$, and having $[LL] = 0$ when L is linear is equivalent to having every pair of operators in L commute. With this as motivation, we borrow some terminology from group theory and say that

Definition 2.7. L is **abelian** if $[LL] = 0$.

We may also take inspiration from ring theory and define *simple* Lie algebras in an analogous way to simple rings. We do just this, but are careful to add an extra condition:

Definition 2.8. A Lie algebra is called **simple** if it has no nontrivial ideals, *and is not abelian.*

There are good reasons for including this latter criterion. One is that it has the effect of immediately excluding any Lie algebra of dimension less than two from being simple, since any such algebra is automatically abelian in view of (L1). We give an example of why this is useful at the end of the next section.
When I is an ideal of L, the quotient space L/I has a well-defined bracket, given by $[x+I, y+I] = [xy]+I$, $x, y \in L$. We call Lie algebras constructed in this way *quotient algebras*, and the surjective homomorphism $\pi : L \to L/I$, $\pi(x) = x + I$ is called the associated *quotient map*. As an application, we observe that $L/[LL]$ is abelian: $[x + [LL], y + [LL]] = [xy] + [LL] = [LL]$, $x, y \in L$.
Let $\phi : L \to M$ be a homomorphism, and define Ker $\phi = \{x \in L \mid \phi(x) = 0\}$. It is an ideal of L, since for any $x \in \text{Ker } \phi$, $\phi([xL]) = [\phi(x)\phi(L)] = [0, \phi(L)] = 0$, meaning $[xL] \subset \text{Ker } \phi$. Similarly, the set $\phi(L) = \{\phi(x) \mid x \in L\}$ is a subalgebra of M. As we will see, ideals of Lie algebras play exactly the same role as ideals in ring theory, in that they bring with them a Lie-algebraic variant of the usual isomorphism theorems. Before we formulate these we list some subalgebras of interest, found in any Lie algebra.

Definition 2.9. Let L be a Lie algebra, X a subset of L, and K a subspace of L (not necessarily a subalgebra). We define

(i) the **centralizer** of X in L to be the set $C_L(X) = \{x \in L \mid [xX] = 0\}$;

(ii) the **center** of L to be the set $Z(L) = \{x \in L \mid [xL] = 0\}$, or equivalently, $Z(L) = C_L(L)$;

(iii) the **normalizer** of K in L to be the set $N_L(K) = \{x \in L \mid [xK] \subset K]\}$.

Verifying that these are subspaces of L is not difficult. To show that they are subalgebras requires the Jacobi identity. Take for example $N_L(K)$. Then

$$[[xy]K] = -[K[xy]] = [x[yK] + [y[Kx]] \subset K + K = K, \quad x, y \in N_L(K),$$

and hence $[xy] \in N_L(K)$. In line with the earlier remark, K is to be understood as being replaced with the same element of K simultaneously across the three leftmost expressions, but to remain as set to the right of the inclusion. That the centralizer is a subalgebra follows similarly. The center, however, is more than a subalgebra; it is an ideal, which follows from $[x, Z(L)] = 0 \subset Z(L)$ for all $x \in L$. If K is a subalgebra, then K is an ideal of $N_L(K)$, which is seen by comparing the definitions. Comparing definitions also reveals that a subspace K is an ideal of L if and only if $N_L(K) = L$.
We now state the isomorphism results we will need thoughout this text. Since they build on already existing isomorphism theorems for vector spaces, their proofs are mostly a matter of verifying that the additional Lie structure is compatible, so we allow ourselves to omit them.

Proposition 2.1. *Let L, M be Lie algebras, $\phi : L \to M$ a homomorphism, and I, J ideals of L. Then*

(a) $L/\mathrm{Ker}\,\phi \cong \phi(L)$;

(b) if $I \subset J$, then J/I is an ideal of L/I, and $(L/I)/(J/I) \cong L/J$;

(c) $(I + J)/J \cong I/(I \cap J)$.

2.2 Modules and representations

Let V be a vector space. Any pair $x \in \mathfrak{gl}(V)$, $v \in V$ yields a vector, namely x evaluated at v. We denote this vector using one of xv, $x(v)$ or $x.v$, depending on the context. The first two will mostly be used when x is fixed, while the third emphasizes evaluation as a function $(x, v) \mapsto x.v$. Since a linear Lie algebra $L \subset \mathfrak{gl}(V)$ consists of linear operators, it is natural to raise the question of in what way the Lie-algebraic structure of L relates to how L maps V into itself, and vice versa. To start with, we can identify three fundamental properties.

(M1) $(ax + by).v = a(x.v) + b(y.v)$;

(M2) $x.(av + bw) = a(x.v) + b(x.w)$;

(M3) $[xy].v = x.(y.v) - y.(x.v)$. $\quad (x, y \in L;\ v, w \in V;\ a, b \in F)$.

Properties (M1) and (M2) follow from L being a linear subspace of $\mathfrak{gl}(V)$ and the elements of L being linear operators respectively, while (M3) is just the definition of the linear bracket.
Much like how abstract Lie algebras generalize linear Lie algebras, we now take these as the defining properties for a type of objects meant to generalize the way a linear Lie algebra acts on its underlying vector space.

Definition 2.10. Let V be a vector space and L a Lie algebra, both over the same field F. Let $f : L \times V \to V$ be a binary operation satisfying (M1)-(M3), which we write $f(x, y) = x.v$. The pair $\langle V, f\rangle$ is then called an **L-module**, and we say that f defines an **action**, or that L **acts** on V.

By abuse of language, we often just say that V is an L-module. Note that while (M3) is equivalent to $[xy].v = (xy).v - (yx).v$ when L is linear and the action is evaluation, we are forced to use expressions such as $x.(y.v)$ in the abstract setting, since in this case xy need not be defined.

Let V be an L-module with action $(x, v) \mapsto x.v$. We say that a subspace W of V is an *L-submodule* of V if $x.w \in W$ for all $x \in L$, $w \in W$. In this case $x.(v + W) = x.w + W$ is well-defined on the quotient space V/W and satisfies (M1)-(M3), so V/W is an L-module. We call such a module a *quotient module*.

Now let V, U be L-modules and let $\phi : V \to U$ be a linear transformation. If ϕ in addition satisfies $\phi(x.v) = x.\phi(v)$ for all $x \in L$, $v \in V$, then we say that ϕ is an *L-module homomorphism*. The kernel of such a homomorphism is an L-submodule of L.

Proposition 2.2. *Let V, U be L-modules and let $\phi : V \to U$ be an L-module homomorphism. Then $V/Ker\ \phi$ is isomorphic as module to $\phi(V)$.*

Given any pair V, W of L-modules, the set $\mathrm{Hom}(V, W)$ of all linear transformations $V \to W$ is itself an L module under the action

$$(x.f)(v) = x.f(v) - f(x.v), \quad x \in L,\ f \in \mathrm{Hom}(V, W),\ v \in V.$$

Note that $x.f = 0$ if and only if f is an L-module homomorphism.

Given $x \in L$ we may define a function $\phi(x) : V \to V$ by letting $\phi(x)(v) = x.v$ for any $v \in V$. Then (M2) guarantees that $\phi(x)$ is linear, so $\phi(x) \in \mathrm{End}\ V$. This yields a function $\phi : L \to \mathfrak{gl}(V)$, which in addition is a linear transformation by (M1). Finally, by (M3),

$$\begin{aligned}\phi([xy])(v) &= [xy].v = x.(y.v) - y.(x.v)\\ &= \phi(x)(\phi(y)(v)) - \phi(y)(\phi(x)(v)) = [\phi(x), \phi(y)](v).\end{aligned}$$

Hence, $\phi : L \to \mathfrak{gl}(V)$ is a homomorphism of Lie algebras. We call a homomorphism of this type a *representation* of L, and our discussion above shows that any L-module induces a representation of L. As might have been guessed this also works in reverse, so that a any homomorphism of the form $\phi : L \to \mathfrak{gl}(V)$ (that is, a representation of L) yields an action by setting $x.v = \phi(x)(v)$. We thus have a bijective correspondence between L-modules and L-representations.

Remark. Let $\phi : L \to \mathfrak{gl}(V)$ be a representation and let W be a submodule of V with respect to the action induced by ϕ. As we have seen, W/V is an L-module, and has a corresponding representation, say $\phi' : L \to \mathfrak{gl}(V/W)$. Explicitly, this representation is given by

$$\phi'(x)(v + W) = x.(v + W) = x.v + W = \phi(x)(v) + W,$$

for all $x \in L$, $v \in V$.

If $L \subset \mathfrak{gl}(V)$ is a linear Lie algebra, then clearly V is an L-module if we take the operator to be evaluation—indeed, this was the motivation for the definition of an L-module. In this case the associated representation is just the inclusion homomorphism of L into $\mathfrak{gl}(V)$. In other words, We will have more to say about representations later. For now we are content with introducing a certain representation central to nearly all of the coming theory. It will play an especially important role in connecting results between linear and abstract Lie algebras.

Definition 2.11. Let K be a subalgebra of L. The action of K on L given by $x.y = [xy]$, $(x \in K,\ y \in L)$ is called the **adjoint action** of K on L, and the representation $K \to \mathfrak{gl}(L)$ induced by this action is called the **adjoint representation** of K in L.

We need to verify that this is an action as claimed. Axioms (M1) and (M2) follow from the bilinearity of the bracket while (M3) follows from the Jacobi identity and anticommutativity:

$$x.(y.z)-y.(x.z) = [x[yz]]-[y[xz]] = [x[yz]+[y[zx]] = -[z[xy]] = [[xy]z] = [xy].z.$$

We use $\mathrm{ad}_L : L \to \mathfrak{gl}(L)$ to denote the adjoint representation in the case where one takes $K = L$. In this notation, a choice of $x \in L$ yields the linear operator $\mathrm{ad}_L\ x \in \mathfrak{gl}(L)$, which, when evaluated at $y \in L$, gives $\mathrm{ad}_L\ x(y) = [xy]$. Again letting K be arbitrary, we see that in this more general case, the corresponding representation is just $\mathrm{ad}_L|_K : K \to \mathfrak{gl}(L)$.
Under the adjoint action, L is a K-module, so we may consider its submodules. These are, by definition, subspaces $I \subset L$ that satisfies $x.y \in I$ for all $x \in K$, $y \in I$. Equivalently, I is a subspace such that $[K, I] \subset I$. When $K = L$, this is just the criterion for I to be an ideal, so the L-submodules of L are exactly the ideals of L. If K is an arbitrary subalgebra, then either of the conditions that I is an ideal L, or $K \subset I$, is sufficient for I to be a K-submodule of L.

Remark. Let $K \subset L$ be a subalgebra, and start with its adjoint representation $\mathrm{ad}_L|_K : K \to \mathfrak{gl}(L)$. Since necessarily $[K, K] \subset K$, we see that K is a K-submodule of L (take $I = K$ in the preceding). Passing to quotients, we obtain the (admittedly rather complicated) representation $(\mathrm{ad}_L|_K)' : K \to \mathfrak{gl}(L/K)$.

We have seen that one translates between actions and representations by setting $\phi(x)(v) = x.v$ in either direction; in the case of the adjoint representation this appears as $\mathrm{ad}_L\ x(y) = [xy]$. In the same way $x \in L$ implies $\phi(x) \in \mathfrak{gl}(L)$ in the general case, we here have $\mathrm{ad}_L\ x \in \mathfrak{gl}(L)$ for $x \in L$. Since representations are homomorphisms we have the identity $[\mathrm{ad}_L\ x, \mathrm{ad}_L\ y] = \mathrm{ad}_L\ [xy]$ for all $x, y \in L$. This can also be calculated directly from the definition $\mathrm{ad}_L\ x(y) = [xy]$. The kernel of ad_L is an ideal and equals

$$\mathrm{Ker}\ \mathrm{ad}_L = \{x \in L \mid \mathrm{ad}_L\ x = 0\} = \{x \in L \mid [xL] = 0\} = Z(L).$$

In other words, any simple Lie algebra is isomorphic to a linear Lie algebra. This would not necessarily be the case if we did not require $[LL] \neq 0$ in our definition of simple algebras, since then $Z(L) = L$ would be possible. This is another indication that our chosen definition is easier to work with.

3 Lie algebras and linear algebra

3.1 The Jordan decomposition

The existence of the adjoint representation seems to suggest that tools from linear algebra could be put to use in probing the structure of abstract Lie algebras. This would be done by applying them to the image of ad_L, and then "pulling back" the results across ad_L. We will accomplish this to various extents, but first we of course need such a set of tools.
Our first definition in this program is rather self-explanatory once we recall some concepts from linear algebra: A linear operator $\mathcal{A}$ is *nilpotent* if there exists a positive integer n such that $\mathcal{A}^n = 0$, and a linear operator over a finite dimensional vector space is *diagonalizable* if its matrix is diagonal for some choice of basis. Rather than using the second term, we will use the term *semisimple* to mean the same thing. There is a technical difference, but this difference vanishes when the field is algebraically closed, which we from now on always assume—this property will also be vital in other stages of the theory.

Definition 3.1. Let L be a Lie algebra and let $x \in L$. We say that x is **ad-nilpotent** if $\mathrm{ad}_L\ x$ is nilpotent. When L is finite dimensional we say that x is **ad-semisimple** if $\mathrm{ad}_L\ x$ semisimple.

This definition raises an immediate question: If L is a (finite dimensional) linear Lie algebra we may speak both about $x \in L$ being nilpotent (diagonalizable) and ad-semisimple (ad-nilpotent)—what is the relationship between these properties, if any? The next lemma provides a partial answer.

Lemma 3.1.

(i) Let L be a linear Lie algebra. If $x \in L$ is nilpotent, then x is ad-nilpotent.

(ii) Let V be a finite dimensional vector space. If $x \in \mathfrak{gl}(V)$ is semisimple, then x is ad-semisimple in $\mathfrak{gl}(\mathfrak{gl}(V))$.

Proof. We start with the nilpotent case. Define $\lambda_x(y) = xy$ and $\rho_x(y) = yx$ for all $y \in L$, so that $\mathrm{ad}_L\ x(y) = [xy] = xy - yx = \lambda_x(y) - \rho_x(y)$, and thus $\mathrm{ad}_L\ x = \lambda_x - \rho_x$. The terms of this difference commute as functions since $\lambda_x(\rho_x(y)) = x(yx) = (xy)x = \rho_x(\lambda_x(y))$. We may therefore use the binomial theorem:

$$(\mathrm{ad}_L\ x)^n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \lambda_x^{n-k} \rho_x^k. \tag{3.1}$$

The assumption that x is nilpotent implies $\lambda_x^m(y) = x^m y = 0$ for some positive integer m, and similarly for ρ_x. Taking n large enough in (3.1) that at least one of the two exponents in each summand is greater than m then forces $(\mathrm{ad}_L\ x)^n = 0$.

Proceeding to the semisimple case, let $n = \dim V$ and pick a basis $(v_1, \ldots, v_n)$ of V in which the matrix of x is diagonal, say $\mathrm{diag}(a_1, \ldots, a_n)$. This choice of basis associates to each element in $\mathfrak{gl}(V)$ a matrix of size $n \times n$. As a vector space, $\mathfrak{gl}(V)$ has dimension n^2 and standard basis $\{e_{ij}\}_{ij}$ $(1 \leq i, j \leq n)$, where e_{ij} is the matrix having 1 at position i, j and 0 everywhere else. We now calculate $\mathrm{ad}_{\mathfrak{gl}(V)}\ x(e_{ij}) = [xe_{ij}] = xe_{ij} - e_{ij}x = a_i e_{ij} - a_j e_{ij} = (a_i - a_j)e_{ij}$. Hence $\mathrm{ad}_{\mathfrak{gl}(V)}\ x$ sends each basis vector of $\mathfrak{gl}(V)$ to a multiple to itself, which is just to say that the matrix of $\mathrm{ad}_{\mathfrak{gl}(V)}\ x$ is diagonal as an operator in $\mathfrak{gl}(\mathfrak{gl}(V))$, i.e. $\mathrm{ad}_{\mathfrak{gl}(V)}\ x$ is semisimple. □

We now formulate a theorem that—roughly speaking—allows us to split an operator into a diagonalizable part and a nilpotent part.

Theorem 3.2 (The Jordan-Chevally decomposition)**.** *Let V be a finite dimensional vector space over an algebraically closed field. For any $\mathcal{A} \in \mathrm{End}\ V$ there exist $\mathcal{A}_s, \mathcal{A}_n \in \mathrm{End}\ V$ such that*

(a) $\mathcal{A} = \mathcal{A}_s + \mathcal{A}_n$; $\mathcal{A}_s$ is diagonalizable, $\mathcal{A}_n$ is nilpotent, $\mathcal{A}_s$ and $\mathcal{A}_n$ commute;

(b) $\mathcal{A}_s$ and $\mathcal{A}_n$ are unique among all pairs of linear operators that satisfy (a);

(c) there exist polynomials $p(t)$ and $q(t)$, both with zero constant term, such that $\mathcal{A}_s = p(\mathcal{A})$ and $\mathcal{A}_n = q(\mathcal{A})$;

(d) for any subspaces $A \subset B \subset V$ such that $\mathcal{A}(B) \subset A$, we also have $\mathcal{A}_s(B) \subset A$ and $\mathcal{A}_n(B) \subset A$.

Proof. See Theorem 8.10 of [6]. □

Remark. A consequence of (b) is that $\mathcal{A}$ is diagonalizable exactly when $\mathcal{A}_n = 0$ in its Jordan decomposition, and similarly $\mathcal{A}_s = 0$ exactly when $\mathcal{A}$ is nilpotent.

In $\mathfrak{gl}(V)$ this decomposition behaves nicely with respect to $\mathrm{ad}_{\mathfrak{gl}(V)}$:

Lemma 3.3. *Let V be finite dimensional. If $x = x_s + x_n$ is the Jordan decomposition of x in $\mathfrak{gl}(V)$, then $\mathrm{ad}_{\mathfrak{gl}(V)}\ x = \mathrm{ad}_{\mathfrak{gl}(V)}\ x_s + \mathrm{ad}_{\mathfrak{gl}(V)}\ x_n$ is the Jordan decomposition of $\mathrm{ad}_{\mathfrak{gl}(V)}\ x$ in $\mathfrak{gl}(\mathfrak{gl}(V))$.*

Proof. First, the adjoints sum up as desired by the linearity of $\mathrm{ad}_{\mathfrak{gl}(V)}$. Secondly, $\mathrm{ad}_{\mathfrak{gl}(V)}\ x_s$ and $\mathrm{ad}_{\mathfrak{gl}(V)}\ x_n$ are, respectively, semisimple and nilpotent in $\mathfrak{gl}(\mathfrak{gl}(V))$ (Lemma 3.1). Lastly, they commute since x_s, x_n do: $[\mathrm{ad}_{\mathfrak{gl}(V)}\ x_s, \mathrm{ad}_{\mathfrak{gl}(V)}\ x_n] = \mathrm{ad}_{\mathfrak{gl}(V)}\ [x_s x_n] = 0$. We need now only invoke the uniqueness of the Jordan decomposition (Theorem 3.2(b)). □

Again using the subscripts s and n to denote the semisimple and nilpotent parts of $\mathrm{ad}_{\mathfrak{gl}(V)}\ x$ in $\mathfrak{gl}(V)$ yields the following concise formulation of Lemma 3.3:

$$(\mathrm{ad}_{\mathfrak{gl}(V)}\ x)_s = \mathrm{ad}_{\mathfrak{gl}(V)}\ x_s, \quad (\mathrm{ad}_{\mathfrak{gl}(V)}\ x)_n = \mathrm{ad}_{\mathfrak{gl}(V)}\ x_n, \quad (x \in \mathfrak{gl}(V)).$$

Note that we may not at this point replace $\mathfrak{gl}(V)$ with an arbitrary linear Lie algebra $L \subset \mathfrak{gl}(V)$ in the above identity; Theorem 3.2 only guarantees, given $x \in L$, the existence of x_s, x_n as operators in $\mathfrak{gl}(V)$, and not that they necessarily lie in the smaller algebra L. We will later find examples of subalgebras for which this stronger property hold.

Remark. Let V be finite dimensional and let $A \subset B \subset \mathfrak{gl}(V)$ be subspaces. Let $x \in \mathfrak{gl}(V)$. Moreover, suppose that $\operatorname{ad}_{\mathfrak{gl}(V)} x(B) \subset A$. Then Theorem 3.2(d) and Lemma 3.3 together imply that

$$\operatorname{ad}_{\mathfrak{gl}(V)} x_s(B) = (\operatorname{ad}_{\mathfrak{gl}(V)} x)_s(B) \subset A.$$

Similarly, $\operatorname{ad}_{\mathfrak{gl}(V)} x_n(B) \subset A$.

3.2 Linear functionals, dual spaces, and bilinear forms

We gather in this section several useful concepts and definitions from linear algebra. These will be used rarely but often decisively throughout the rest of the text. Some calculations have been formulated as remarks in order to highlight their importance.

Let V be a vector space over F. Given subsets $S \subset V$ and $A \subset F$ we define the *A-span* of S to be the set of all linear combinations of vectors in S with coefficients in A. Clearly the A-span of S is contained in V, and the F-span of S is just the usual span of S in V and therefore a vector subspace of V. Now let U be another vector space over some field E, and let $B \subset F \cap E$ be a subset. We say that the function $f : V \to U$ is *B-linear* if it satisfies $f(av + bw) = af(v) + bf(w)$ for all $v, w \in V$, $a, b \in B$. If $F = E = B$ then f is B-linear if and only if f is a linear transformation.

We can view F itself as a vector space over F by simply taking the field operations as the vector space operations. It is one-dimensional since F is the F-span of any nonzero $a \in F$. Similarly, if K is any subfield of F then F is a vector space over K. We use F_K to represent this point of view. The subspaces of F_K are exactly the K-span of subsets of F. Unlike F as a vector space over itself F_K need not in general be one-dimensional.

A *linear functional* on V is a linear transformation from V to F when we view the latter as a vector space. The *dual space* of V, which we denote V^*, is the vector space of all linear functionals on V. If V is finite dimensional with basis $(v_1, \dots, v_n)$ then V^* has a corresponding basis $(f_1, \dots, f_n)$, where f_i is the linear functional defined by $f_i(v_j) = \delta_{ij}$ (the Kronecker delta). Hence V^* is finite dimensional, and $\dim V^* = \dim V$. Moreover, $\sum_i a_i v_i \mapsto \sum_i a_i f_i$ is an isomorphism of vector spaces.

Remark. Let V be finite dimensional over F. Let K be a subfield of F, let E be a subspace of F_K, and let $f \in E^*$. Suppose that in some basis, $x, y \in \mathfrak{gl}(V)$ has matrices $\operatorname{diag}(a_1, \dots, a_m)$ and $\operatorname{diag}(f(a_1), \dots, f(a_m))$ respectively, where $a_1, \dots, a_m \in E$. Let $\{e_{ij}\}_{ij}$ be the associated basis of $\mathfrak{gl}(V)$. Of the m^2 pairs $\big(a_i - a_j, f(a_i) - f(a_j)\big)$, $1 \le i, j \le m$ some may have their first components equal, but then so are their second components: $a_i - a_j = a_k - a_l$ implies $f(a_i) - f(a_j) = f(a_k) - f(a_l)$ by the linearity of f. The set of pairs therefore associates to every unique first component a unique second component. We may now construct their Lagrange polynomial $p(t)$, a polynomial with no constant term satisfying $p(a_i - a_j) = f(a_i) - f(a_j)$ for all $1 \le i, j \le m$. Recall the proof of Lemma 3.1, which when repeated here shows that $\operatorname{ad}_{\mathfrak{gl}(V)} x(e_{ij}) = (a_i - a_j)e_{ij}$

and $\mathrm{ad}_{\mathfrak{gl}(V)}\ y(e_{ij}) = (f(a_i) - f(a_j))e_{ij}$. Then, for all $1 \leq i, j \leq m$,

$$p(\mathrm{ad}_{\mathfrak{gl}(V)}\ x)(e_{ij}) = p(a_i - a_j)e_{ij} = \mathrm{ad}_{\mathfrak{gl}(V)}\ y(e_{ij}).$$

Writing $y = f(x)$ we have the formula $\mathrm{ad}_{\mathfrak{gl}(V)}\ f(x) = p(\mathrm{ad}_{\mathfrak{gl}(V)}\ x)$.

Any field with characteristic 0 has a subfield isomorphic to the field of rational numbers. This subfield, which we by abuse of language denote Q, is generated by adjoining the identity in F to Q and then closing Q under addition, additive inverses, multiplication, and multiplicative inverses. We may therefore always form the vector space F_Q as long as char $F = 0$.

Remark. Given a finite subset $\{a_1, \ldots, a_m\} \subset F$, how may we approach showing that every element in it is identically zero? Take E to be the Q-span of $\{a_1, \ldots, a_m\}$, so E is a subspace of F_Q. We want to show that $E = 0$, and for this it suffices to show that $E^* = 0$. Hence let $f \in E^*$, that is, $f : E \to Q$ and f is Q-linear. Suppose we knew that $\sum_i a_i f(a_i) = 0$. Apply f to both sides and use the Q-linearity of f to get $\sum_i f(a_i)^2 = 0$. A sum of squares of rational numbers is zero if and only if every rational number in the sum is zero, so $f(a_i) = 0$ $(1 \leq i \leq m)$. Then f must be zero on the Q-span of $\{a_1, \ldots, a_m\}$ and this span is by definition E, so $f = 0$. We therefore see that a sufficient condition for all a_i to be zero is for every $f \in E^*$ to satisfy $\sum_i a_i f(a_i) = 0$.

Let V be finite dimensional. The *trace* of any linear operator $\mathcal{A} \in \mathrm{End}\ V$ is defined as the sum of the diagonal elements of the matrix of $\mathcal{A}$, in some basis of V. This sum—which we denote $\mathrm{tr}(\mathcal{A})$—is invariant under change of basis, and therefore well-defined. When F is algebraically closed we may equivalently define $\mathrm{tr}(\mathcal{A})$ as the sum of the eigenvalues of $\mathcal{A}$, counted with multiplicity. Hence $\mathrm{tr}(\mathcal{A}) = 0$ if $\mathcal{A}$ is nilpotent, since the unique eigenvalue of $\mathcal{A}$ is 0. Observe that the function $\mathrm{tr} : \mathrm{End}\ V \to F$ given by $\mathcal{A} \mapsto \mathrm{tr}(\mathcal{A})$ is a linear functional since

$$\mathrm{tr}(a\mathcal{A} + b\mathcal{B}) = a\ \mathrm{tr}(\mathcal{A}) + b\ \mathrm{tr}(\mathcal{B}), \quad \mathcal{A}, \mathcal{B} \in \mathrm{End}\ V,\ a, b \in F. \tag{3.2}$$

In other words, $\mathrm{tr} \in (\mathrm{End}\ V)^*$. The trace also satisfies the useful identity

$$\mathrm{tr}(\mathcal{A}\mathcal{B}) = \mathrm{tr}(\mathcal{B}\mathcal{A}), \quad \mathcal{A}, \mathcal{B} \in \mathrm{End}\ V. \tag{3.3}$$

Remark. Let L be a finite dimensional linear Lie algebra. Then the trace of any operator of L is defined. In this case we have an additional identity:

$$\mathrm{tr}([xy]z) = \mathrm{tr}(x[yz]), \quad x, y, z \in L. \tag{3.4}$$

We refer to this as the trace being *associative* on L. To see that (3.4) holds, expand the brackets on each side and use linearity together with (3.3).

A *bilinear form* on V is a bilinear function $\beta : V \times V \to F$. We say that β is *symmetric* if $\beta(v, w) = \beta(w, v)$ for all $v, w \in V$. If β is a bilinear form over a Lie algebra L we say that β is *associative* if $\beta([xy], z) = \beta(x, [yz])$, $x, y, z \in L$. Now let V be finite dimensional. From (3.2) and (3.3) we see that the function $(x, y) \mapsto \mathrm{tr}(xy)$ is a symmetric bilinear form on V, called the *trace form* of V, which we write as $\mathrm{Tr}(x, y)$ for $x, y \in \mathrm{End}\, V$. The trace form is defined on any finite dimensional linear Lie algebra, so in particular $\mathfrak{gl}(V)$. It is associative by (3.4), i.e. $\mathrm{Tr}([xy], z) = \mathrm{Tr}(x, [yz])$ for all $x, y, z \in \mathrm{End}\, V$.

Let V be a vector space (not necessarily finite dimensional), W a subspace of V, and β a bilinear form on V. The *orthogonal complement* of W in V is

$$W^{\perp} = \{v \in V \mid \beta(v, W) = 0\}.$$

The bilinearity of β implies that $W^{\perp}$ is a subspace of V. We give a special name to $V^{\perp}$, which we call the *radical* of β, and usually denote S. If the radical is of β is zero then we say that β is *nondegenerate*. If V is finite dimensional and β is nondegenerate then $\dim W + \dim W^{\perp} = \dim V$. In this case $W = V$ is equivalent to $W^{\perp} = 0$.

Remark. Let L be a Lie algebra and let β be an associative bilinear form on L. If I is an ideal of L then so is $I^{\perp}$: For any $x, y \in I^{\perp}$ we have $[xy] \in I^{\perp}$, since

$$\beta([xy], I) = \beta(x, [yI]) \subset \beta(x, I) = 0.$$

A special case is $L^{\perp}$, since L is an ideal of itself. The radical S is hence an ideal.

Any bilinear form β on a vector space V yields a linear transformation from V into its dual V^*. This linear transformation is furnished by the mapping $v \mapsto (w \mapsto \beta(v, w))$. Call this mapping ϕ; in other words $\phi(v)$, $v \in V$ is the linear functional $\phi(v) : V \to F$ defined by $\phi(v)(w) = \beta(v, w)$. That this is indeed a linear functional follows from the linearity of the second argument of β, and that ϕ is a linear transformation follows from the linearity of the first. The kernel of ϕ is exactly the radical of β, so β is nondegenerate if and only if ϕ is injective. Now if V is finite dimensional then we know from earlier that $\dim V = \dim V^*$. In other words *any nondegenerate bilinear form on a finite dimensional vector space yields a natural isomorphism between the vector space and its dual.* The adjective "natural" here refers to the fact that we could construct this isomorphism without choosing a basis of V, i.e. without making any "unnatural" or arbitrary choices. Under these assumptions we see that to every $f \in V^*$ is associated a unique element $v_f \in V$ such that $\phi(v_f) = f$, or equivalently $\beta(v_f, v) = f(v)$ for all $v \in V$. A basis $(f_1, \ldots, f_n)$ of V^* therefore yields a unique basis $(v_1^*, \ldots, v_n^*)$ of V satisfying $\beta(v_i^*, v) = f_i(v)$ for all $v \in V$, $i = 1, \ldots, n$.

Remark. If we in the above nevertheless choose a basis $(v_1, \ldots, v_n)$ of V, then as before V^* has a corresponding basis $(f_1, \ldots, f_n)$, where f_i is the linear functional defined by $f_i(v_j) = \delta_{ij}$. Then there exists a unique basis $(v_1^*, \ldots, v_n^*)$ of V satisfying $\beta(v_i^*, v_j) = \delta_{ij}$ for $i, j = 1, \ldots, n$. This basis $(v_i^*)_i$ is called the *dual basis* of $(v_i)_i$. Now let L be a finite dimensional Lie algebra with basis $(x_1, \ldots, x_n)$, and let β be a nondegenerate associative bilinear form on L. Let $(y_1, \ldots, y_n)$ be the dual basis, which satisfies $\beta(y_i, x_j) = \delta_{ij}$. For any $x \in L$ there are coefficients $a_{ij}, b_{ij} \in F$, $i, j = 1, \ldots, n$ such that

$$[xx_i] = \textstyle\sum_j a_{ij} x_j, \quad [xy_i] = \textstyle\sum_j b_{ij} y_j.$$

We may calculate how these coefficients relate to each other as follows:

$$\begin{aligned} a_{ik} &= \sum_j a_{ij}\delta_{kj} = \sum_j a_{ij}\beta(y_k, x_j) = \beta(y_k, [xx_i]) \\ &= \beta([y_k x], x_i) = -\sum_j b_{kj}\beta(y_j, x_i) = -\sum_j b_{kj}\delta_{ji} = -b_{ki}. \end{aligned} \tag{3.5}$$

Now suppose L is linear. We may then also calculate $[x, x_i y_i]$ using (3.5) and the identity $[x, yz] = [xy]z + y[xz]$, $x, y, z \in \mathfrak{gl}(V)$:

$$[x, x_i y_i] = [xx_i]y_i + x_i[xy_i] = \textstyle\sum_j a_{ij}x_i y_i - \sum_j a_{ji}x_i y_i.$$

As a consequence, $[x, \sum_i x_i y_i] = \sum_i [x, x_i y_i] = 0$. This holds for any $x \in L$; hence $\sum_i x_i y_i$ commutes with L, or $\sum_i x_i y_i \in Z(L)$.

4 Special classes of Lie algebras

4.1 Nilpotent algebras

The concept of ad-nilpotency alone does not take us very far in our program beyond the results we already have—we need yet another notion of nilpotency in our treatment of abstract Lie algebras. To motivate such a definition, let L be linear and let every element in L be nilpotent. We might conceivably call such an algebra "nilpotent" in itself. Now, if $x \in L$ then x is ad-nilpotent by Lemma 3.1(i), so there exists a positive integer n such that

$$\underbrace{[x[x \dots [x}_{n}\, y]]] = (\mathrm{ad}_L\, x)^n(y) = 0, \quad (y \in L).$$

One way to generalize this feature is to demand the existence of a *single* positive integer n such that

$$[x_1[x_2 \dots [x_n y]]] = 0, \quad (x_1, \dots, x_n, y \in L). \tag{4.1}$$

With this in mind we give our next definition.

Definition 4.1. Let L be a Lie algebra. Define $L^0 = L$ and $L^i = [LL^{i-1}]$ for $i = 1, 2, \dots$. We say that L is **nilpotent** (as Lie algebra) if there exists a positive integer n such that $L^n = 0$.

Recall that if I, J are ideals of L then $[IJ]$ is defined to be the ideal of all finite linear combinations of elements of the form $[xy]$, $x \in I$, $y \in J$. By induction, $L^0, L^1, L^2, \dots$ are ideals of L. Moreover, $L^0 \supset L^1 \supset L^2 \supset \dots$, also by induction. We call this descending chain of ideals the *lower central series* of L. Hence L is nilpotent if and only if its lower central series terminates; that is, every ideal is zero beyond some point in the chain.
Let L be nilpotent. By (4.1), every element of L is ad-nilpotent. Our main theorem of this section says that the converse of this also is true, given that L is finite dimensional.

Theorem 4.1 (Engel)**.** *Let L be a finite dimensional Lie algebra. If every element of L is ad-nilpotent, then L is nilpotent.*

We sometimes have the chance to apply this theorem on linear Lie algebras, in which case we may use a streamlined version.

Corollary 4.1.1. *Let L be a finite dimensional linear Lie algebra. If every element of L is nilpotent, then L is nilpotent.*

Proof. Use Lemma 3.1(i) and Engel's Theorem. □

Before we can prove Engel's Theorem, we need some basic facts about nilpotent Lie algebras, and another theorem.

Proposition 4.2. *Let L be a Lie algebra. Then*

(a) if L is nilpotent, then all subalgebras and homomorphic images of L are;

(b) if $L/Z(L)$ is nilpotent, then L is nilpotent;

(c) if L is nilpotent and $L \neq 0$, then $Z(L) \neq 0$.

Proof. (a) Let K be a subalgebra of L. Assume that $K^i \subset L^i$, which clearly holds for $i = 0$. Then $K^{i+1} = [KK^i] \subset [LL^i] = L^{i+1}$, so $K^i \subset L^i$ by induction on i. Similarly, if $\phi : L \to M$ is a surjective homomorphism, then induction shows that $M^i = \phi(L^i)$, and therefore the lower central series of K and M terminates whenever the lower central series of L does.

(b) Let $\pi : L \to L/Z(L)$ be the quotient homomorphism with kernel $Z(L)$. By the proof of part (a), $(L/Z(L))^i = \pi(L)^i = \pi(L^i)$, so $\pi(L^n) = 0$ for some positive integer n. Then L^n lies in the kernel of π, that is, $L^n \subset Z(L)$, so $L^{n+1} = [LL^n] \subset [LZ(L)] = 0$ by the definition of $Z(L)$.

(c) Let n be the unique positive integer such that $L^n \neq 0$ but $L^{n+1} = 0$. Then $L^n \subset Z(L)$ by $[LL^n] = 0$ and the definition of $Z(L)$.

□

We have put the theorem needed to prove Engel's Theorem in the appendix due to its proof being slightly longer while at the same time not venturing much beyond the techniques already seen before this section (one does not require the above proposition, for example). The curious reader can find the proof in Section 8.1. In any case, the theorem states that *if $L \subset \mathfrak{gl}(V)$, $(V \neq 0)$ is a Lie algebra of nilpotent operators, then there exists nonzero $v \in V$ such that $L.v = 0$, where the action is evaluation.*
We are now ready to prove Engel's Theorem.

Proof of Theorem 4.1. If $L = 0$ then there is nothing to prove, so suppose L is nonzero. Take as induction assumption that the theorem holds for any Lie algebra of dimension less than L. By the conditions of the theorem every element in $\text{ad}_L\ L$ is nilpotent in $\mathfrak{gl}(L)$, so if we take L as our vector space and $\text{ad}_L\ L$

as our Lie algebra, then these satisfy the conditions of the mentioned theorem. This shows that there exists some nonzero $y \in L$ such that $\mathrm{ad}_L\ x(y) = 0$ for all $x \in L$, or equivalently, $[Ly] = 0$. Then $y \in Z(L)$, so that $Z(L) \neq 0$ which shows that $L/Z(L)$ has dimension less than L. Furthermore, every element $x + Z(L) \in L/Z(L)$ is ad-nilpotent since for some n depending on x, $\left(\mathrm{ad}_{L/Z(L)}\ (x + Z(L))\right)^n (y + Z(L)) = (\mathrm{ad}_L\ x)^n(y) + Z(L) = 0 + Z(L) = Z(L)$ for all $y + Z(L) \in L/Z(L)$. We may then apply our induction assumption to get that $L/Z(L)$ is nilpotent; hence L is nilpotent by Proposition 4.2(b). $\square$

We end this section with a useful nilpotency criterion for operators in $\mathfrak{gl}(V)$.

Lemma 4.3. *Let V be finite dimensional and let $A \subset B \subset \mathfrak{gl}(V)$ be subspaces. Define $M = \{x \in \mathfrak{gl}(V) \mid [xB] \subset A\}$. A sufficient condition for $x \in M$ to be nilpotent is to satisfy* $\mathrm{Tr}(x, M) = 0$.

Proof. Let $x = x_s + x_n$ be the Jordan decomposition of x in $\mathfrak{gl}(V)$. The hypothesis $x \in M$ is, by the definition of M, equivalent to having $\mathrm{ad}_{\mathfrak{gl}(V)}\ x(B) \subset A$. Then $\mathrm{ad}_{\mathfrak{gl}(V)}\ x_s(B) \subset A$ by the final remark of Section 3.1, so $x_s \in M$. Recall the remark immediately after Theorem 3.2, which says that x is nilpotent if and only if $x_s = 0$. Now x_s is semisimple, so in some basis its matrix is diagonal, say $\mathrm{diag}(a_1, \ldots, a_m)$. Put $E = \mathrm{span}_Q\{a_1, \ldots, a_m\}$ and let $f \in E^*$. According to one the remarks in Section 3.2 we are done if we can show that $\sum_i a_i f(a_i) = 0$, since then $E = 0$ and $x_s = 0$. Taking $K = Q$ in an earlier remark of the same section furnishes a polynomial $p(t)$ without constant term such that $\mathrm{ad}_{\mathfrak{gl}(V)}\ f(x_s) = p(\mathrm{ad}_{\mathfrak{gl}(V)}\ x_s)$. If an operator maps a subspace B into a subspace A, then so does any polynomial expression without constant term of that operator, so $\mathrm{ad}_{\mathfrak{gl}(V)}\ x_s(B) \subset A$ implies that $\mathrm{ad}_{\mathfrak{gl}(V)}\ f(x_s)(B) \subset A$. Hence $f(x_s) \in M$. By hypothesis $\mathrm{tr}(xf(x_s)) = 0$, but then

$$0 = \mathrm{tr}(xf(x_s)) = \mathrm{tr}(x_s f(x_s)) + \mathrm{tr}(x_n f(x_s)) = \mathrm{tr}(x_s f(x_s)) = \textstyle\sum_i a_i f(a_i).$$

This is due to the appearance of the matrix of x_n in our chosen basis: Looking at the proof of the existence of the Jordan decomposition, one sees that it has zeros everywhere except possibly on the "diagonal" immediately above the main diagonal, where it may also have ones. Then it commutes with $f(x_s)$, so $x_n f(x_s)$ is nilpotent and $\mathrm{tr}(x_n f(x_s)) = 0$. $\square$

Let M be as in the lemma. If we take $M^\perp$ with respect to the trace form on $\mathfrak{gl}(V)$ (that is, $M^\perp = \{x \in \mathfrak{gl}(V) \mid \mathrm{Tr}(x, M) = 0\}$), then the lemma just says that all $x \in M \cap M^\perp$ are nilpotent.

4.2 Solvable algebras

Our next family of Lie algebras are called *solvable* algebras, and their definition closely mirrors that of nilpotent algebras.

Definition 4.2. Given a Lie algebra L, set $L^{(0)} = L$ and $L^{(i)} = [L^{(i-1)} L^{(i-1)}]$ for $i = 1, 2, \ldots$. We say that L is **solvable** if there exists a positive integer n such that $L^{(n)} = 0$.

Our first connection between nilpotent and solvable algebras other than the similarity of the definitions is that, using induction, $L^{(i)} = [L^{(i-1)}L^{(i-1)}] \subset [LL^{i-1}] = L^i$, with base case $L^{(0)} = L = L^0$. This shows that nilpotent algebras are solvable. Nilpotency is in fact (as might be expected) a strictly stronger property, i.e. there are solvable algebras that are not nilpotent, though we will refrain from calculating any examples until later. As a second connection, observe that $L^{(n)} \subset [LL]^{n-1}$ when $n = 1$ (they are equal), and if this is true for $n = k$ then

$$L^{(k+1)} = [L^{(k)}L^{(k)}] \subset [L[LL]^{k-1}] = [LL]^k.$$

By induction we have $L^{(n)} \subset [LL]^{n-1}$ for all positive integers n; hence L is solvable if $[LL]$ is nilpotent. We call the sequence $L^{(0)}, L^{(1)}, L^{(2)}, \ldots$ the *derived series* of L, and as in the nilpotent case these are all ideals of L satisfying $L^{(0)} \supset L^{(1)} \supset L^{(2)} \supset \ldots$. Some other useful properties of solvable algebras are gathered below. Note the similarities with Proposition 4.2.

Proposition 4.4. *Suppose L is a Lie algebra. Then*

(a) if L is solvable, then all subalgebras and homomorphic images of L are

(b) if I is an ideal of L such that I and L/I are solvable, then L is solvable;

(c) if I and J are solvable ideals of L, then $I + J$ is solvable.

Proof. (a) We have $K^{(i)} \subset L^{(i)}$ when K is a subalgebra of L, which can be verified in the same way as the proof of Theorem 4.2(a). If $\phi : L \to M$ is a surjective homomorphism, take as induction assumption that $M^{(i)} = \phi(L^{(i)})$ for some $i = 0, 1, 2 \ldots$ (it is clearly true for $i = 0$). Then $M^{(i+1)} = [M^{(i)}M^{(i)}] = [\phi(L^{(i)})\phi(L^{(i)})] = \phi([L^{(i)}L^{(i)}]) = \phi(L^{(i+1)})$, so by induction $M^{(i)} = \phi(L^{(i)})$ for all $i = 0, 1, 2, \ldots$. This shows that if the derived series of L terminates, then so does the derived series of K and M.

(b) Let $\pi : L \to L/I$ be the quotient homomorphism with kernel I. By the proof of part (a), $(L/I)^{(i)} = \pi(L)^{(i)} = \pi(L^{(i)})$, so $\pi(L^{(n)}) = 0$ for some positive integer n. Then $L^{(n)}$ lies in the kernel of π, that is, $L^n \subset I$. Also, $I^{(m)} = 0$ for some positive integer m since I is solvable. Then $L^{(n+m)} = \left(L^{(n)}\right)^{(m)} \subset I^{(m)} = 0$. That we can rewrite $L^{(n+m)}$ in this way follows from the identity $L^{(i+1)} = [L^{(i)}L^{(i)}] = \left(L^{(i)}\right)^{(1)}$ and induction.

(c) Note first that $I/(I \cap J)$ is solvable by part (a) since it is the homomorphic image of I under the quotient homomorphism, and I is solvable. Then $(I + J)/J$ is solvable since it is isomorphic to $I/(I \cap J)$ by Proposition 2.1. Now apply part (b) to see that $I + J$ is solvable.

□

There is a theorem similar to the one used in proving Engel's Theorem, in that it, too, guarantees the existence of vectors acted upon in a certain way, but this time for a solvable linear Lie algebra. We refer the reader to Section 8.1 for a

complete formulation. We will make use of this theorem only first in Section 4.4, so one can postpone knowing it until then.
Engel's Theorem gave us a criterion for nilpotency; now we derive a criterion for solvability.

Theorem 4.5 (Cartan's Criterion). *Let V be finite dimensional and let L be a subalgebra of $\mathfrak{gl}(V)$. A sufficient condition for L to be solvable is to satisfy* $\mathrm{Tr}([LL], L) = 0$.

Proof. Write $A = [LL]$ and $B = L$; clearly $A \subset B \subset \mathfrak{gl}(V)$ and are subspaces (since they are subalgebras). Now define M as in Lemma 4.3:

$$M = \{x \in \mathfrak{gl}(V) \mid [xB] \subset A\} = \{x \in \mathfrak{gl}(V) \mid [xL] \subset [LL]\}.$$

The lemma says that a sufficient condition for $x \in M$ to be nilpotent is to satisfy $\mathrm{Tr}(x, M) = 0$. We have $[LL] \subset L \subset M$ by the definition of M, which shows that if $\mathrm{Tr}([LL], M) = 0$ then every element of $[LL]$ is nilpotent, and $[LL]$ is in turn nilpotent as algebra (Corollary 4.1.1). This is not our hypothesis, but does in fact follow from it. To see this, observe that any $x \in [LL]$ may (by construction of $[LL]$) be written of the form $x = \sum_{i=1}^{n} a_i[y_i z_i]$ for some $y_i, z_i \in L$, $a_i \in F$, and n a positive integer. Then, using that $[xM] \subset [LL]$ for all $x \in L$,

$$\begin{aligned} \mathrm{Tr}(x, M) &= \sum_{i=1}^{n} a_i \mathrm{Tr}([y_i z_i], M) = \sum_{i=1}^{n} a_i \mathrm{Tr}(y_i, [z_i M]) \\ &\subset \sum_{i=1}^{n} a_i \mathrm{Tr}(y_i, [LL]) \subset \mathrm{Tr}(L, [LL]) = \mathrm{Tr}([LL], L) = 0. \end{aligned}$$

Hence $[LL]$ is nilpotent and L solvable (as shown in the beginning of the section). □

Corollary 4.5.1. *Let L be a finite dimensional Lie algebra. A sufficient condition for L to be solvable is to satisfy* $\mathrm{Tr}(\mathrm{ad}_L\ [LL], \mathrm{ad}_L\ L) = 0$.

Proof. Take $V = L$ in Cartan's Criterion with $\mathrm{ad}_L\ L \subset \mathfrak{gl}(L)$ as the subalgebra. The hypothesis of the criterion is $\mathrm{Tr}([\mathrm{ad}_L\ L, \mathrm{ad}_L\ L], \mathrm{ad}_L\ L) = 0$, but this is exactly the hypothesis of the corollary since $[\mathrm{ad}_L\ x, \mathrm{ad}_L\ y] = \mathrm{ad}_L\ [xy]$ for all $x, y \in L$. Hence $\mathrm{ad}_L\ L = \mathrm{Im}\ \mathrm{ad}_L \cong L/\mathrm{Ker}\ \mathrm{ad}_L = L/Z(L)$ is solvable. But $Z(L)$ is always solvable, so L is solvable by Proposition 4.4(b). □

4.3 The Killing form and semisimple algebras

All our considerations up to this point seem to invite us to take the "preimage" of the trace form on $\mathrm{ad}_L\ L$ in order to obtain a bilinear form on L. This we do.

Definition 4.3. Let L be a finite dimensional Lie algebra. Its **Killing form** κ is the bilinear form defined by $\kappa(x, y) = \mathrm{Tr}(\mathrm{ad}_L\ x, \mathrm{ad}_L\ y),\ x, y \in L$.

Bilinearity, symmetry, and associativity follow from those of Tr. The first and last of these also requires the homomorphism properties of ad_L. Restating Corollary 4.5.1 in terms of this form tells us that a sufficient condition for L to be solvable is to satisfy $\kappa([LL], L) = 0$.
The Killing form of a proper subalgebra K of L need not in general be the restriction $\kappa|_{K\times K}$, since $\mathrm{ad}_K\, x$ and $\mathrm{ad}_L\, x$ are different as operators (if nothing else their corresponding matrices have different sizes, regardless of basis). When K is an ideal, however, the Killing forms of K and L does in fact coincide.

Lemma 4.6. *Let I be an ideal of L. If κ_I and κ_L are their respective Killing forms, then $\kappa_I = \kappa_L|_{I\times I}$.*

Proof. Pick a basis of I and extend it to a basis of L, so that an element of I in column vector form has zeroes for every entry with index greater than $\dim I$. Fix $x, y \in I$, and set $\mathcal{A} = \mathrm{ad}_L\, x\, \mathrm{ad}_L\, y$. We calculate that $\mathcal{A}(L) = [x[yL]] \subset [xI] \subset I$. Identify $\mathcal{A}$ with its matrix in the aforementioned base; it must have zeroes on all rows with index greater than $\dim I$, since otherwise $\mathcal{A}$ would send some elements of L outside I. Also, the submatrix consisting of the first $\dim I$ rows and columns of $\mathcal{A}$ is exactly the matrix corresponding to $\mathrm{ad}_I\, x\, \mathrm{ad}_I\, y$ in this base. Taken together this shows that $\mathrm{tr}(\mathcal{A}) = \mathrm{tr}(\mathrm{ad}_I\, x\, \mathrm{ad}_I\, y)$, or in other words, $\kappa_L(x, y) = \kappa_I(x, y)$. □

As a first application, we observe that any simple Lie algebra has nondegenerate Killing form. Recall from Section 3.2 that this is to say that the *radical* belonging to the form, or

$$S = \{x \in L \mid \kappa(x, L) = 0\},$$

is zero. To see that this is indeed the case when L is simple, use that S is an ideal, so one of $S = L$, $S = 0$ is true. If it were the former, then $\kappa(L, L) = 0$, meaning L is solvable. But this contradicts $[LL] = L$, which is necessarily the case when L is simple (since otherwise $[LL]$ would either be zero or a nonzero ideal of L, and both alternatives are impossible by assumption).
An interesting question is to what extent the converse of this holds: If L has nondegenerate Killing form, what can we say about its ideals? Before we investigate this we need another definition.

Definition 4.4. The unique maximal solvable ideal of L is called the **radical** of L and is denoted Rad L.

Maximal here means not included in any larger solvable ideal. A maximal solvable ideal certainly has to exist (since L is finite dimensional), but its uniqueness is less obvious. To prove it, recall Proposition 4.4(c), which states that $I + J$ is a solvable ideal whenever I and J are. Then J being maximal would imply $I + J = J$, that is, $I \subset J$. The same holds for I, so any two maximal solvable ideals contain each other and thus must be equal.
It may seem as if we have overloaded the word "radical", since we already use it in conjunction with bilinear forms—however, this is not without good reason.

By defintion the Killing form of L is nondegenerate if and only if its radical S is zero. Now compare to the next theorem.

Theorem 4.7. *The Killing form of L is nondegenerate if and only if Rad $L = 0$.*

Proof. Write $R = \text{Rad } L$. First suppose $S = 0$. If $R \neq 0$ then R furnishes a nonzero solvable abelian ideal; for this, take the ideal $R^{(n)}$ where n is the smallest positive integer such that $R^{(n+1)} = 0$, which is solvable and abelian since $[R^{(n)}R^{(n)}] = R^{(n+1)}$. If we can show that all abelian ideals of L are zero, then Rad $L = 0$ too. Thus let I be an abelian ideal, and let $x \in I$. Given any $y \in L$, put $\mathcal{A} = \text{ad}_L\ x\ \text{ad}_L\ y$. We have $\mathcal{A}^2 z = [x[y[x[yz]]] \in [I[yI]] \subset [II] = 0$ for all $z \in L$, so $\mathcal{A}$ is nilpotent. Then it has vanishing trace, or $0 = \text{tr}(\mathcal{A}) = \kappa(x, y)$. This holds for all $y \in L$, so $x \in S$, which shows that $I \subset S = 0$.

For the reverse direction, suppose that Rad $L = 0$. Given any $x \in [SS]$ and $y \in S$ we have $\text{tr}(\text{ad}_S\ x\ \text{ad}_S\ y) = \kappa(x, y) = 0$ by Lemma 4.6 and $x \in S$. As remarked earlier, $\kappa([SS], S) = 0$ is sufficient for S to be solvable, and this is exactly what we have showed here. Being a solvable ideal of L, S must be contained in Rad $L = 0$; hence $S = 0$. □

To return to the question of how the nondegeneracy of the Killing form affects the ideal structure of L, we will soon see that under this condition, L is roughly speaking built up using simple ideals for the atomic components. It is then natural to name such algebras *semisimple*. We employ Theorem 4.7 to slightly reformulate this definition.

Definition 4.5. We say that L is **semisimple** if Rad $L = 0$.

We prefer this definition since it implies that $L/\text{Rad } L$ is semisimple even when L is not, which is in agreement with our intuition of quotient algebras as "dividing out" an ideal along with some characterizing property of the ideal. To see this, let M be the preimage of Rad $(L/\text{Rad } L)$ along the quotient homomorphism $\pi : L \to L/\text{Rad } L$. Then $\pi(M^{(n)}) = (\pi(M))^{(n)} = (\text{Rad } (L/\text{Rad } L))^{(n)} = 0$ for some n, which gives $M^{(n)} \subset \text{Ker } \pi = \text{Rad } L$. Hence M is solvable and $M \subset \text{Rad } L$. But $\text{Rad } L \subset M$ by construction, so $M = \text{Rad } L$. Therefore we have Rad $(L/\text{Rad } L) = 0$.

Every semisimple Lie algebra is perforce finite dimensional, since otherwise its radical is not guaranteed to exist. As such we will from now on omit specifying that a given Lie algebra is finite dimensional if we have already assumed it to be semisimple.

4.4 Structural considerations

Let V be a finite-dimensional vector space, and let $n = \dim n$. A *flag* in V is a chain of subspaces $0 = V_0 \subset V_1 \subset \cdots \subset V_n = V$, where the dimension of each subspace is one greater than the one before it, or $\dim V_i = i$ for all i. As long as V is nonzero, we may take quotients and obtain a flag

$$0 = V_1/V_1 \subset V_2/V_1 \subset \cdots \subset V_n/V_1 = V/V_1 \tag{4.2}$$

in V/V_1. This also works in the opposite direction: A flag in V/Fv, ($v \in V$ nonzero) necessarily has to be of the form (4.2) for $V_1 = Fv$ and some subspaces V_i, $i = 2, 3, \ldots, n$, due to that part of the isomorphism theorems for vector spaces that relate the subspace structure of V to that of its quotients. Clearly $V_i \subset V_{i+1}$, $(1 \leq i \leq n-1)$, and it follows that $0 = V_0 \subset V_1 \subset \cdots \subset V_n = V$ is a flag in V by dimensional considerations. We now set out to show that

Proposition 4.8. *Let $L \subset \mathfrak{gl}(V)$ be a linear Lie algebra, where V is finite-dimensional. Write $n = \dim V$.*

(a) If every operator in L is nilpotent, then there exists a flag $\{V_i\}_i$ in V such that $L.V_i \subset V_{i-1}$ for all $i = 1, \ldots, n$.

(b) If L is solvable, then there exists a flag $\{V_i\}_i$ in V such that $L.V_i \subset V_i$ for all $i = 1, \ldots, n$.

Proof. The case $n = 0$ is trivial. If $n = 1$, then the flag can just be taken to be $0 = V_0 \subset V_n = V$—(b) is clear, and a nilpotent operator of a one-dimensional space (which hence acts as a scalar) is necessarily the zero operator, so (a) is also satisfied. We use induction for the remaining cases.

(a) Theorem 8.1 furnishes a nonzero vector $v \in V$ such that $L.v = 0$. Set $V_1 = Fv$ (the F-span of $\{v\}$). The action here is evaluation and the corresponding representation is just the inclusion map $i : L \to \mathfrak{gl}(V)$. The space V_1 is killed by L and hence a submodule, so we have a quotient representation $i' : L \to \mathfrak{gl}(V/V_1)$ defined by $i'(x).(w + V_1) = x.w + V_1$. With the vector space being V/V_1, the algebra $M = i'(L)$ is seen to satisfy the hypotheses of (a), since $i'(x)$ is nilpotent if x is by the preceeding formula. Also, $\dim V/V_1 = n - 1$, so induction yields a flag $\{V_i/V_1\}_i$ in V/V_1, which immediately translates into a flag $\{V_i\}_i$ in V. It remains to show that $L.V_i \subset V_{i-1}$ for $i = 2, \ldots, n$ (the case $i = 1$ is already known from $L.v = 0$). We know from induction that $M.(V_i/V_1) \subset V_{i-1}/V_1$, $(2 \leq i \leq n)$. But

$$\begin{aligned} M.(V_i/V_1) &= \{i'(x).(w + V_1) \mid x \in L, w \in V_i\} \\ &= \{x.w + V_1 \mid x \in L, w \in V_i\} = (L.V_i)/V_1, \end{aligned}$$

so it follows that $L.V_i \subset V_{i-1}$, $(2 \leq i \leq n)$.

(b) Theorem 8.2 provides a nonzero vector $v \in L$ along with a linear functional $\lambda \in L^*$ such that $x.v = \lambda(x)v$ for all $x \in L$. We do not actually need λ in its full specificity; the fact that it implies $L.V_1 \subset V_1$ is enough. After constructing V_1 and M as in (a), we argue similarly and obtain a flag $\{V_i\}_i$ in V (here we have to use that the image of a solvable algebra is solvable). Again it only remains to show that $L.V_i \subset V_i$, $(2 \leq i \leq n)$, the proof of which is entirely analogous to that of (a).

□

Remark. Part (b) of the proposition is usually cited as Lie's Theorem.

If $L \neq 0$ is solvable (nilpotent) and we choose a basis of V in such a way that each basis element hails from exactly one of the subspaces of the flag furnished by the proposition, then the extra condition on the flag says that relative to this basis, the matrices corresponding to the elements of L are upper triangular (strictly upper triangular). For convenience, let $\mathfrak{t}(n, F)$ and $\mathfrak{n}(n, F)$ denote, respectively, the set of $n \times n$ upper triangular and strict upper triangular matrices having elements in F. (It is easy to verify that these are linear Lie algebras and that $[\mathfrak{t}(n, F), \mathfrak{t}(n, F)] \subset \mathfrak{n}(n, F)$.) We will only need the full power of the proposition in order to prove the next corollary, which itself will only see a single application in Section 6.2. Recall the beginning of Section 4.2, in which we computed that L is solvable if $[LL]$ is nilpotent. With Lie's Theorem, we are now in a position where we can obtain the equivalency of these statements.

Corollary 4.8.1. *If L is a finite-dimensional solvable Lie algebra, then L has a flag of ideals, and $[LL]$ is nilpotent.*

Proof. Let $n = \dim L$. Consider $\mathrm{ad}_L\ L \subset \mathfrak{gl}(L)$; it is the homomorphic image of a solvable subalgebra and hence solvable by Proposition 4.4(a). It is also linear, so we may use Lie's Theorem to find a flag $\{L_i\}_i$ in L which furthermore satisfies $(\mathrm{ad}_L\ L).L_i \subset L_i$, $(1 \leq i \leq n)$. In other words, $[LL_i] \subset L_i$ for all i, and we have the first part of the corollary. For the second, observe that it suffices to prove that all $\mathrm{ad}_{[LL]}\ x$, $(x \in [LL])$ are nilpotent for $[LL]$ to be, according to Engel's Theorem. But $\mathrm{ad}_{[LL]}\ x = (\mathrm{ad}_L\ x)|_{[LL]}$, $(x \in [LL])$, so we can just work with the $\mathrm{ad}_L\ x$, $(x \in [LL])$ instead. To show that these are nilpotent, use that $\mathrm{ad}_L\ L$ being solvable implies that it has a base relative to which its elements are upper triangular, i.e. lie in $\mathfrak{t}(2, F)$. Then

$$\mathrm{ad}_L\ [LL] = [\mathrm{ad}_L\ L, \mathrm{ad}_L\ L] \subset [\mathfrak{t}(n, F), \mathfrak{t}(n, F)] \subset \mathfrak{n}(2, F),$$

where we by the first inclusion mean more of an identification. The rightmost algebra certainly consists of nilpotent operators, and we are done. □

We now turn our attention to semisimple algebras. Let L semisimple and let $I_1, \ldots, I_n$ be ideals of L. We say that L is the *direct sum* of these ideals given that it is their direct sum as subspaces, and we write $L = I_1 \oplus \cdots \oplus I_n$. For this to be the case the ideals must be disjoint, and as a consequence $[I_i I_j] \subset I_i \cap I_j = 0$, $(i \neq j)$. Hence we have $[\sum_i x_i, \sum_i y_i] = \sum_i [x_i y_i]$, $(x_i, y_i \in I_i,\ 1 \leq i \leq n)$. This says that applying the bracket in L is equivalent to "componentwise" evaluation of the brackets of I_i, in the sense that $x_i \in I_i$ is one component of $x = \sum_i x_i$. This notation is used in the next theorem, which provides a comprehensive answer to our earlier question about what consequences a nondegenerate Killing form has for the ideal structure of L.

Theorem 4.9. *Let L be semisimple with Killing form κ. Then L possesses simple ideals $L_1, \ldots, L_n$ such that*

(i) $L = L_1 \oplus \cdots \oplus L_n$;

(ii) $\kappa_{L_i} = \kappa|_{L_i \times L_i}$ *for all* $i = 1, \ldots, n$;

(iii) if I is any simple ideal of L, then $I = L_i$ for some $i = 1, \ldots, n$.

Proof. We first show (i) through induction on $\dim L$. If L has no proper nonzero ideals, then L is simple and trivially satisfies the theorem, so assume otherwise. Let L_1 be a minimal nonzero ideal of L. Then $L_1^\perp$ (the orthogonal subspace of L_1 with respect to the Killing form) is an ideal, and so is $J = L_1 \cap L_1^\perp$. We have $\kappa([JJ], J) = 0$ by the definition of $L_1^\perp$, which as seen earlier is sufficient for J to be solvable, so $J = 0$ by the semisimplicity of L. The sum $L_1 + L_1^\perp$ is therefore direct and has dimension equal to $\dim L_1 + \dim L_1^\perp$. This dimension is equal to $\dim L$ by the nondegeneracy of κ, which shows that $L = L_1 \oplus L_1^\perp$.

Now let I be any ideal of L_1. We have $[LI] = [L_1 I] \oplus [L_1^\perp I] = I \oplus 0 = I$, so I is also an ideal of L. We chose L_1 to be minimal, so I must be trivial. Thus L_1 is simple. Any ideal of $L_1^\perp$ is similarly an ideal of L. In particular, any solvable ideal of L_1 or $L_1^\perp$ is a solvable ideal of L, so L being semisimple implies that L_1 and $L_1^\perp$ are semisimple. Hence $L_1^\perp$ satisfies the induction hypothesis and can be written as a direct sum $L_1^\perp = L_2 \oplus \cdots \oplus L_n$ of simple ideals of $L_1^\perp$. Then $L = L_1 \oplus L_2 \oplus \cdots \oplus L_n$ where each L_i is a simple ideal of L.

Part (ii) of the theorem now follows from (i) and Lemma 4.6.

Let I be any simple ideal of L. Observe that $[IL] = [IL_1] \oplus \cdots \oplus [IL_n]$. Each of $[IL], [IL_1], \ldots, [IL_n]$ is an ideal of I and so must be either zero or I. Recall that $Z(L)$ is solvable and hence zero, i.e. $[xL] = 0$, $x \in L$ implies $x = 0$. The simple ideal I is a fortiori nonzero, which forces $[IL] = I$ by the above. This is only possible if $[IL_i] = I$ for exactly one $i = 1, \ldots, n$; say $i = k$. Then $I = [IL_k]$ is a nonzero ideal of the simple ideal L_k, so $I = L_k$ and we are done. □

We have seen that subalgebras and homomorphic images of nilpotent (solvable) algebras are themselves nilpotent (solvable). As might be expected, a similar result holds for semisimple algebras.

Corollary 4.9.1. *If L is semisimple, then*

(a) each ideal of L is a direct sum of simple ideals of L;

(b) all ideals and homomorphic images of L are semisimple;

(c) $L = [LL]$.

Proof. Write L as a direct sum of simple ideals $L_1, \ldots, L_n$ as in Theorem 4.9. The first part of the proof of the theorem is valid when L_1 is replaced with an arbitrary proper ideal I. More specifically, any ideal of I is an ideal of L, I is semisimple, and I decomposes (by induction) into simple ideals, so then $I = L_{i_1} \oplus \cdots \oplus L_{i_k}$ for some $i_1, \ldots, i_k$. This proves (a) and the first part of (b). As for (c), componentwise application of the bracket gives

$$\begin{aligned}[LL] &= [L_1 \oplus \cdots \oplus L_n, L_1 \oplus \cdots \oplus L_n] \\ &= [L_1 L_1] \oplus \cdots \oplus [L_n L_n] \\ &= L_1 \oplus \cdots \oplus L_n \\ &= L.\end{aligned}$$

For the other part of (b), let $\phi : L \to M$ be a surjective homomorphism. Then $M \cong L/\text{Ker}\ \phi \cong (L_1 \oplus \cdots \oplus L_n)/(L_{i_1} \oplus \cdots \oplus L_{i_k}) \cong L_{j_1} \oplus \cdots \oplus L_{j_l}$. The converse of Theorem 4.9 is certainly true (Rad L is solvable, but also a direct sum of simple and therefore nonsolvable ideals of L, and this is only possible if Rad $L = 0$), so M is semisimple. □

Note also part (c), which implies that the family of nilpotent and/or solvable algebras and the family of semisimple algebras are disjoint.

5 Consequences of semisimplicity

5.1 Representations of semisimple algebras

Consider the adjoint representation $\text{ad}_L : L \to \mathfrak{gl}(L)$, or equivalently the action $x.y = [xy]$. When L is semisimple, Theorem 4.9 gives L as a direct sum $L = L_1 \oplus \cdots \oplus L_n$, where each L_i is an ideal and hence an L-submodule of L. Moreover, L_i is simple and so has no nontrivial L-submodules. Since the bracket evaluates componentwise, the action of L on itself is determined in full by the actions of L_i on themselves. Motivated by this, we introduce some definitions.

Definition 5.1. An L-module is *irreducible* if it is nonzero and its only submodules are the trivial ones. A finite dimensional L-module is **completely reducible** if it is a direct sum of irreducible L-modules. If ϕ is a representation of L whose corresponding L-module is irreducible (completely reducible) then we say that ϕ also has this property.

Note that any one-dimensional module necessarily is irreducible. Our aim in this section is to generalize the discussion preceeding the definition by proving that, in fact, *every* finite dimensional representation of a semisimple algebra is completely reducible. We start with the following lemma, which, roughly speaking, generalizes the triviality of the adjoint action of a one-dimensional algebra on itself.

Lemma 5.1. *Let L be semisimple and $\phi : L \to \mathfrak{gl}(V)$ a finite dimensional representation of L. If W is any one-dimensional submodule of V, then L acts trivially on W, i.e. $L.W = 0$.*

Proof. Let $\rho : L \to \mathfrak{gl}(W)$ be the corresponding representation. The operators in $\mathfrak{gl}(W)$ act as scalars since W is one-dimensional, but Corollary 4.9.1(c) gives $\text{tr}(\rho(L)) = \text{tr}(\rho([LL])) = \text{tr}([\rho(L)\rho(L)]) = 0$, which forces $\rho(L) = 0$. □

Before proceeding we need to develop some theory relating to representations. This comes in the form of the so-called *Casimir element.*
Let $\phi : L \to \mathfrak{gl}(V)$ be a finite dimensional representation. The Killing form is the special case $V = L$, $\phi = \text{ad}_L$ of the following construction: Define a symmetric associative bilinear form β on L by $\beta(x, y) = \text{Tr}(\phi(x), \phi(y))$, $x, y \in L$.

If in addition L is semisimple and ϕ is *faithful*, i.e. injective, then β is also nondegenerate. To see this, let S be the radical of β. Then

$$\mathrm{Tr}(\phi(S), \phi(S)) = \beta(S, S) = 0,$$

so $\phi(S)$ is solvable by Cartan's Criterion. But $S \cong \phi(S)$ by faithfulness, so S is a solvable ideal of L and hence zero by semisimplicity.
Continuing, let L be semisimple, ϕ faithful, and define β as above. Fix a basis $(x_i)_i$ of L. Since β is nondegenerate and associative, we may construct the dual basis $(y_i)_i$ as outlined at the end of Section 3.2. Clearly $(\phi(x_i))_i$ and $(\phi(y_i))_i$ are bases of $\phi(L)$ by faithfulness. Moreover, $(\phi(y_i))_i$ is the dual base of $(\phi(x_i))_i$ with respect to the trace form of $\phi(L)$, since

$$\mathrm{Tr}(\phi(y_i), \phi(x_j)) = \beta(y_i, x_j) = \delta_{ij}.$$

Now, $\phi(L)$ is a linear Lie algebra, so by the final remark of Section 3.2 we have

$$\sum_i \phi(x_i)\phi(y_i) \in Z(\phi(L)) = C_{\mathfrak{gl}(V)}(\phi(L)). \tag{5.1}$$

Write $\sum_i \phi(x_i)\phi(y_i) = c_\phi(\beta)$; we call this the *Casimir element of* β, and it commutes with $\phi(L)$ by (5.1). The trace of $c_\phi(\beta)$ is easily calculated:

$$\mathrm{tr}(c_\phi(\beta)) = \sum_i \mathrm{tr}(\phi(x_i)\phi(y_i)) = \sum_i \beta(x_i, y_i) = \sum_i \beta(y_i, x_i) = \sum_i \delta_{ii} = \dim L.$$

More can be said in the case when the representation is irreducible by applying Schur's Lemma (the proof of which we omit).

Lemma 5.2 (Schur)**.** *Let $\phi : L \to \mathfrak{gl}(V)$ be an irreducible representation. Then $x \in C_{\mathfrak{gl}(V)}(\phi(L))$ if and only if x acts as a scalar on V.*

Supposing that ϕ is irreducible, the lemma shows that the only possibility is $c_\phi(\beta) = \frac{\dim L}{\dim V} I_V$, since the trace evaluates incorrectly otherwise. In this special case, we see that the exact form of $c_\phi(\beta)$ is independent of the choice of basis (though one can verify that this is also true for arbitrary representations).
It is possible to obtain an element having the properties of the Casimir element even when ϕ is not faithful. This is done in the following way: Ker ϕ is an ideal of L and hence a direct sum of simple ideals of L. Take L' to be the direct sum of the remaining ideals; L' is semisimple and the restriction $\phi|_{L'}$ is a faithful representation of L', so we may construct the Casimir element $c = c_{\phi|_{L'}}(\beta) \in \mathfrak{gl}(V)$. It commutes with $\phi(L')$, but clearly $\phi(L) = \phi(L')$, so c commutes with $\phi(L)$ (and as before, $\mathrm{tr}(c) = \dim L$). We then (by abuse of language) call c the Casimir element of ϕ.
We are now ready to prove the earlier mentioned result on complete reducibility, though rather than using Definition 5.1, we instead verify an equivalent statement: ϕ is completely reducible if and only if each to each submodule $W \subset V$ belongs a submodule $W' \subset V$ such that $V = W \oplus W'$—a *complement* of W.

Theorem 5.3 (Weyl)**.** *If L is semisimple, then any finite dimensional representation $\phi : L \to \mathfrak{gl}(V)$ is completely reducible.*

Proof. Let W be a proper nonzero submodule of V. Our strategy will be to show that $V = W \oplus \text{Ker } f$ for an appropriate choice of L-homomorphism $f : V \to W$. One can show that $\text{Hom}(V, W) = \{f : V \to W \mid f \text{ is linear}\}$ is an L-module under the action that maps $(x, f) \in L \times \text{Hom}(V, W)$ to the linear transformation $x.f \in \text{Hom}(V, W)$ defined by $(x.f)(v) = x.f(v) - f(x.v)$, $v \in V$. Notice that $L.f = 0$ if and only if f is an L-homomorphism. Let $\mathscr{V} \subset \text{Hom}(V, W)$ consist of those f for which $f|_W$ acts as a scalar—say $f|_W = a_f I_W$, $a_f \in F$. Let $\mathscr{W} \subset \mathscr{V}$ consist of those f for which $f|_W = 0$ ($a_f = 0$). Now, given $(x, f) \in L \times \mathscr{V}$, we calculate that $(x.f)|_W = 0$ as follows: For any $w \in W$,

$$(x.f)(w) = x.f(w) - f(x.w) = x.f|_W(w) - f|_W(x.w) = a_f x.v - a_f x.v = 0.$$

In other words, $L.\mathscr{V} \subset \mathscr{W}$, so $\mathscr{W} \subset \mathscr{V} \subset \text{Hom}(V, W)$ are in fact submodules. Hence, the quotient module $\mathscr{V}/\mathscr{W}$ is defined. Also, if $f + \mathscr{W} \neq \mathscr{W}$ ($a_f \neq 0$), then $g + \mathscr{W} = \frac{a_g}{a_f}(f + \mathscr{W})$ for any $g \in \mathscr{V}$, which shows that $\mathscr{V}/\mathscr{W}$ is one-dimensional. Next, we find a complement of $\mathscr{W}$ in $\mathscr{V}$. There are two cases to consider:

(1) $\mathscr{W}$ is irreducible.

Let $\rho : L \to \mathfrak{gl}(\mathscr{V})$ be the representation induced by the L-module $\mathscr{V}$, and let c be the Casimir element of ρ. It commutes with $\rho(L)$, so for all $x \in L$, $f \in \mathscr{V}$ we have

$$0 = [c, \rho(x)](f) = c(\rho(x)(f)) - \rho(x)(c(f)) = c(x.f) - x.c(f).$$

Hence $c : \mathscr{V} \to \mathscr{V}$ is an L-module homomorphism. Consequently, $X = \text{Ker } c$ is a submodule of $\mathscr{V}$. The module $\mathscr{V}/\mathscr{W}$ is one-dimensional, so $L.(\mathscr{V}/\mathscr{W}) = \mathscr{W}$ (Lemma 5.1), or equivalently $\rho(L)(\mathscr{V}) \subset \mathscr{W}$. Also, c is a sum of compositions of elements of $\rho(L)$, which means that $c(\mathscr{V}) \subset \mathscr{W}$. By the rank-nullity theorem, X is one-dimensional. Fix a basis of $\mathscr{W}$ and extend it to a basis of $\mathscr{V}$—as in the proof of Lemma 4.6, we obtain $\text{tr}(c|_{\mathscr{W}}) = \text{tr}(c) = \dim L$. Evidently, $c|_{\mathscr{W}}$ commutes with the image of the irreducible representation $L \to \mathfrak{gl}(\mathscr{W})$ induced by $\mathscr{W}$ (this image is just $\rho(L)|_{\mathscr{W}}$), so Shur's Lemma together with the preceding implies that $c|_{\mathscr{W}} = \frac{\dim L}{\dim \mathscr{W}} I_{\mathscr{W}}$. Therefore $\mathscr{W} \cap X = 0$, and it follows by dimensional considerations that $\mathscr{V} = \mathscr{W} \oplus X$.

(2) $\mathscr{W}$ is reducible.

Let $\mathscr{W}'$ be a proper nonzero L-submodule of $\mathscr{W}$. Consider the following diagram.

$$0 \longrightarrow \mathscr{W} \longrightarrow \mathscr{V} \longrightarrow \mathscr{V}/\mathscr{W} \longrightarrow 0$$

The maps are, from left to right, the trivial, inclusion, projection, and zero map. They form an exact sequence of linear transformations (i.e. the image of each is the kernel of the next). The leftmost three are L-module homomorphisms, and Lemma 5.1 implies that the last is, too—the diagram is therefore an *exact sequence of L-module homomorphisms.* As in the theory of "ordinary" modules, we obtain a new exact sequence by passing to quotient modules:

$$0 \longrightarrow \mathscr{W}/\mathscr{W}' \longrightarrow \mathscr{V}/\mathscr{W}' \longrightarrow \tfrac{\mathscr{V}}{\mathscr{W}'}/\tfrac{\mathscr{W}}{\mathscr{W}'} \longrightarrow 0.$$

Of the nonzero modules, note that the two leftmost ones have lower dimension than those of the previous sequence, while the rightmost is isomorphic to $\mathscr{V}/\mathscr{W}$ and still one-dimensional. Introduce the induction hypothesis that there exists a one-dimensional submodule $\tilde{\mathscr{W}}/\mathscr{W}' \subset \mathscr{V}/\mathscr{W}'$ such that

$$\tfrac{\mathscr{V}}{\mathscr{W}'} = \tfrac{\mathscr{W}}{\mathscr{W}'} \oplus \tfrac{\tilde{\mathscr{W}}}{\mathscr{W}'} \tag{5.2}$$

(to motivate this hypothesis, observe that we can repeat the quotient process until the leftmost module is irreducible, in which case (1) furnishes such a submodule). Writing out the associated maps of the quotient module $\tilde{\mathscr{W}}/\mathscr{W}'$ gives the exact sequence

$$0 \longrightarrow \mathscr{W}' \longrightarrow \tilde{\mathscr{W}} \longrightarrow \tilde{\mathscr{W}}/\mathscr{W}' \longrightarrow 0.$$

Again, by induction, there exists a one-dimensional submodule $X \subset \tilde{\mathscr{W}}$ such that $\tilde{\mathscr{W}} = \mathscr{W}' \oplus X$. Being one-dimensional, $\mathscr{W} \cap X \neq 0$ would imply $X \subset \mathscr{W}$. By the direct sum (5.2) we have $(\mathscr{W} \cap \tilde{\mathscr{W}})/\mathscr{W}' = (\mathscr{W}/\mathscr{W}') \cap (\tilde{\mathscr{W}}/\mathscr{W}') = 0$, that is, $\mathscr{W} \cap \tilde{\mathscr{W}} \subset \mathscr{W}'$, and it would follow that $X \subset \mathscr{W} \cap X \subset \mathscr{W} \cap \tilde{\mathscr{W}} \subset \mathscr{W}'$. But this would contradict $\mathscr{W}' \cap X = 0$, so we must have $\mathscr{W} \cap X = 0$. As in (1) we have $\dim \mathscr{W} + \dim X = (\dim \mathscr{V} - 1) + 1 = \dim V$, and therefore $\mathscr{V} = \mathscr{W} \oplus X$.

From (1) and (2) we have a one-dimensional submodule $X \subset \mathscr{V}$ such that $\mathscr{V} = \mathscr{W} \oplus X$. Take any nonzero $f \in X$. It spans X and necessarily satisfies $a_f \neq 0$, since otherwise $f \in \mathscr{W}$, which would contradict the direct sum. We may thus assume $f|_W = I_W$—if not, replace f with f/a_f. The action of L on X is trivial (Lemma 5.1). In particular, $x.f = 0$, so f is an L-homomorphism. Hence, f is a left split in the exact sequence

$$0 \longrightarrow W \underset{f}{\overset{i}{\rightleftarrows}} V \longrightarrow V/W \longrightarrow 0,$$

where i is the inclusion map, since evidently $f \circ i = f|_W = I_W$. By the splitting lemma for L-modules, $V = \operatorname{Im} i \oplus \operatorname{Ker} f = W \oplus \operatorname{Ker} f$. Since it is a submodule, $\operatorname{Ker} f$ is therefore our sought complement to W. □

5.2 Derivations

Given an F-algebra U we may consider those linear operators $\mathcal{A} \in \operatorname{End} U$ that satisfy the so-called product rule:

$$\mathcal{A}(x \cdot y) = x \cdot \mathcal{A}y + \mathcal{A}x \cdot y, \quad (x, y \in U). \tag{5.3}$$

For the sake of clarity we here use $\cdot$ instead of juxtaposition to write the bilinear operation of the F-algebra. Operators that satisfy (5.3) are called *derivations.* One example is provided by the infinite-dimensional vector space $\mathbb{R}[x]$ of real polynomials, which is an $\mathbb{R}$-algebra under polynomial multiplication, and with the usual derivative $p(x) \mapsto \frac{d}{dx}p(x)$ as one possible derivation. The set of all derivations of U is denoted Der U, and for the derivations themselves we use the greek letters δ, ϵ, etc.

Remark. In a Lie algebra L over F—which, in particular, is an F-algebra with the bracket for the bilinear operator—the product rule takes the form

$$\delta([xy]) = [x, \delta(y)] + [\delta(x), y], \quad (\delta \in \text{Der } L,\ x, y \in L).$$

Of course, if L happens to be linear, then we may view it as an F-algebra in a different way, with composition of operators as the bilinear operation instead. In this case the product rule appears as $\delta(xy) = x\delta(y)+\delta(x)y$. It is straightforward to verify that a derivation in this latter sense is also a derivation in the former. When we speak of Der L, we will always mean the set of derivations with respect to the bracket, regardless of whether L is linear or not.

The set Der U is a subspace of End U, which we see by showing that $a\delta + b\epsilon$ satisfies the product rule for any $\delta, \epsilon \in \text{Der } U$; $a, b \in F$.

$$\begin{aligned}(a\delta + b\epsilon)(x \cdot y) &= a\delta(x \cdot y) + b\epsilon(x \cdot y)\\ &= a(x \cdot \delta(y) + \delta(x) \cdot y) + b(x \cdot \epsilon(y) + \epsilon(x) \cdot y)\\ &= x \cdot (a\delta(y) + b\epsilon(y)) + (a\delta(x) + b\epsilon(x)) \cdot y\\ &= x \cdot (a\delta + b\epsilon)(y) + (a\delta + b\epsilon)(x) \cdot y, \quad (x, y \in U)\end{aligned}$$

In fact, Der U is a subalgebra of $\mathfrak{gl}(U)$, since the bracket of any two derivations is again a derivation:

$$\begin{aligned}[\delta, \epsilon](x \cdot y) = (\delta\epsilon - \epsilon\delta)(x \cdot y) &= \delta(x \cdot \epsilon(y) + \epsilon(x) \cdot y) - \epsilon(x \cdot \delta(y) + \delta(x) \cdot y)\\ &= x \cdot \delta(\epsilon(y)) + \delta(x) \cdot \epsilon(y) + \epsilon(x) \cdot \delta(y) + \delta(\epsilon(x)) \cdot y\\ &\qquad - x \cdot \epsilon(\delta(y)) - \epsilon(x) \cdot \delta(y) - \delta(x) \cdot \epsilon(y) - \epsilon(\delta(x)) \cdot y\\ &= x \cdot \delta(\epsilon(y)) + \delta(\epsilon(x)) \cdot y - x \cdot \epsilon(\delta(y)) - \epsilon(\delta(x)) \cdot y\\ &= x \cdot (\delta\epsilon - \epsilon\delta)(y) + (\delta\epsilon - \epsilon\delta)(x) \cdot y\\ &= x \cdot [\delta, \epsilon](y) + [\delta, \epsilon](x) \cdot y, \quad (x, y \in U).\end{aligned}$$

Observe that $\delta\epsilon$ (as well as $\epsilon\delta$) denotes the composition of δ and ϵ as linear operators and not, say, the function $x \mapsto \delta(x) \cdot \epsilon(x)$, which need not even be linear (as is the case if we take $\delta = \epsilon = \frac{d}{dx}$ in our example with the $\mathbb{R}$-algebra of real polynomials).

Now let L be a Lie algebra. To see why derivations will play an important role in understanding the structure of L, we note that not only is it true that $\text{ad}_L\ L \subset \text{Der } L$—$\text{ad}_L\ L$ is in fact an ideal of Der L. The first of these claims can be verified by using the Jacobi identity. To prove the second, let $\delta \in \text{Der } L$, $x \in L$. Then

$$[\delta, \text{ad}_L\ x](y) = \delta([xy]) - [x, \delta(y)] = [\delta(x), y] = \text{ad}_L\ \delta(x)(y), \quad (y \in L). \tag{5.4}$$

In other words, $[\delta, \mathrm{ad}_L\ L] \subset \mathrm{ad}_L\ L$ for all $\delta \in \mathrm{Der}\ L$. When L is semisimple we can say even more:

Theorem 5.4. *If L is semisimple, then* $\mathrm{ad}_L\ L = \mathrm{Der}\ L$.

Proof. Write $M = \mathrm{ad}_L\ L$, $D = \mathrm{Der}\ L$. By the preceding, M is an ideal of D, and hence $M^\perp = \{x \in D \mid \kappa_D(x, M) = 0\}$ is, too. Being an ideal, M has non-degenerate Killing form (Theorem 4.7), and this form satisfies $\kappa_M = \kappa_D|_{M\times M}$ (Lemma 4.6). All in all, this shows that

$$\begin{aligned}[M^\perp, M] &\subset M^\perp \cap M \\ &= \{x \in M \mid \kappa_D(x, M) = 0\} \\ &= \{x \in M \mid \kappa_M(x, M) = 0\} = 0.\end{aligned}$$

Now let $\delta \in M^\perp$ be arbitrary. By the above and (5.4), we see that $\mathrm{ad}_L\ (\delta(x)) = [\delta, \mathrm{ad}_L\ x] = 0$ for all $x \in L$. Semisimplicity forces $Z(L) = 0$, so $\mathrm{ad}_L : L \to M$ is an isomorphism (in particular, injective). Therefore $\delta(x) = 0$ for all $x \in L$, i.e. $\delta = 0$. This proves that $M^\perp = 0$, which is equivalent to $M = D$. □

Together with the next lemma, this theorem provides a way to extend the Jordan decomposition to abstract semisimple Lie algebras. We introduce the method for this in the next section. In proving the lemma, we will make use of the following terminology: A *chain* in V is a set of nonzero vectors

$$\mathcal{A}^{k-1}v, \quad \mathcal{A}^{k-2}v, \quad \ldots, \quad \mathcal{A}v, \quad v,$$

for some $\mathcal{A} \in \mathrm{End}\ V$, $v \in V$, where k is a positive integer, with the additional condition that $\mathcal{A}^k v = 0$. Note also that while the two theorems cited in the proof each are formulated for the case when F is the field of complex numbers, their proofs still go through when F is an algebraically closed field of characteristic zero (this is even explicitly mentioned in the first source).

Lemma 5.5. *Let U be a finite dimensional F-algebra. Then* $\mathrm{Der}\ U$ *is closed under the Jordan decomposition, i.e. if $\delta \in \mathrm{Der}\ U$ has Jordan decomposition $\delta = \delta_s + \delta_n$ in* $\mathrm{End}\ U$*, then $\delta_s, \delta_n \in \mathrm{Der}\ U$.*

Proof. Let $\delta = \delta_s + \delta_n$ be as in the lemma, and let E denote the set of all (distinct) eigenvalues of δ_s. According to Theorem 7.8 of [6], U has a basis in which the basis vectors are organized in chains of the form

$$\begin{matrix} (\delta - \lambda_1 I)^{k_1 - 1}x_1, & \ldots, & (\delta - \lambda_1 I)x_1, & x_1, \\ \vdots & \vdots & \vdots & \vdots \\ (\delta - \lambda_m I)^{k_m - 1}x_m, & \ldots, & (\delta - \lambda_m I)x_m, & x_m, \end{matrix}$$

where $\lambda_i \in E$ for each i. The eigenvalues corresponding to different chains need not be unique, i.e. it is possible that $\lambda_i = \lambda_j$ when $i \neq j$. Most importantly, by construction of the Jordan decomposition, δ_s *is diagonal relative to this basis, with matrix*

$$\operatorname{diag}(\underbrace{\lambda_1,\ldots,\lambda_1}_{k_1},\ldots,\underbrace{\lambda_m,\ldots,\lambda_m}_{k_m}).$$

Another construction vital to this proof is the collection of subspaces

$$U_a = \{x \in U \mid (\delta - aI)^k x = 0 \text{ for some positive integer } k = k(x)\},$$

where $a \in F$. We write $k(x)$ to emphasize that k, if it exists, may be different for different $x \in U$. According to Theorem 8.23 of [1], U decomposes as

$$U = \bigoplus_{\lambda \in E} U_\lambda. \tag{5.5}$$

Evidently, the basis chains corresponding to the eigenvalue λ all lie in U_λ, and from the aforementioned appearance of δ_s relative to our chosen basis, we see that $\delta_s(x) = \lambda x$ for all $x \in U_\lambda$. Together, these facts provide us with a complete description of how δ_s acts on U.

Before proceeding, we need some information about how the bilinear operation behaves with respect to these subspaces. Given any $a, b \in F$, set $\mathcal{A} = \delta - aI$, $\mathcal{B} = \delta - bI$, $\mathcal{C} = \delta - (a+b)I$. We start by computing the identity

$$\begin{aligned} \mathcal{C}(x \cdot y) &= \delta(x \cdot y) - (a+b)(x \cdot y) \\ &= \delta(x) \cdot y - (ax) \cdot y + x \cdot \delta(y) - x \cdot (by) \\ &= \mathcal{A}x \cdot y + x \cdot \mathcal{B}y. \end{aligned} \tag{5.6}$$

Now consider the following statement.

$$\mathcal{C}^n(x \cdot y) = \sum_{i=0}^{n} \binom{n}{i} \mathcal{A}^{n-i}x \cdot \mathcal{B}^i y, \quad (n \geq 0) \tag{5.7}$$

It certainly holds for $n = 0$ since both sides reduce to $x \cdot y$. Supposing that (5.7) holds for some non-negative integer n, we may apply (5.6) together with Pascal's identity in order to verify the statement for $n+1$:

$$\begin{aligned} \mathcal{C}^{n+1}(x \cdot y) &= \mathcal{C}^n(\mathcal{A}x \cdot y) + \mathcal{C}^n(x \cdot \mathcal{B}y) \\ &= \sum_{i=0}^{n} \binom{n}{i} \mathcal{A}^{n+1-i}x \cdot \mathcal{B}^i y + \sum_{i=0}^{n} \binom{n}{i} \mathcal{A}^{n-i}x \cdot \mathcal{B}^{i+1} y \\ &= \sum_{i=1}^{n} \left[\binom{n}{i} + \binom{n}{i-1} \right] \mathcal{A}^{n+1-i}x \cdot \mathcal{B}^i y \\ &\qquad + \mathcal{A}^{n+1}x \cdot y + x \cdot \mathcal{B}^{n+1} y \\ &= \sum_{i=1}^{n} \binom{n+1}{i} \mathcal{A}^{n+1-i}x \cdot \mathcal{B}^i y + \mathcal{A}^{n+1}x \cdot y + x \cdot \mathcal{B}^{n+1} y \\ &= \sum_{i=0}^{n+1} \binom{n+1}{i} \mathcal{A}^{n+1-i}x \cdot \mathcal{B}^i y. \end{aligned}$$

Hence (5.7) holds for all $n \geq 0$ by induction. Choosing n large enough in (5.7) yields $C^n(x \cdot y) = 0$ for any given $x \in U_a$, $y \in U_b$, which means that $x \cdot y \in U_{a+b}$. In other words, $U_a \cdot U_b \subset U_{a+b}$ for all $a, b \in F$.

Now suppose $x, y \in U$ are nonzero. By (5.5), there exist $a, b \in E$ such that $x \in U_a$, $y \in U_b$. Additionally, $x \cdot y \in U_{a+b}$ by the preceding. This subspace is nonempty if $a + b \in E$ and empty if not; in either case, we have $\delta_s(x \cdot y) = (a + b)x \cdot y$. But $\delta_s(x) \cdot y + x \cdot \delta_s(y) = (a + b)x \cdot y$, so δ_s satisfies the product rule: $\delta_s(x \cdot y) = \delta_s(x) \cdot y + x \cdot \delta_s(y)$. Of course, this rule also holds whenever one of x, y is zero. We conclude that $\delta_s \in \text{Der } U$, which also implies that $\delta_n = \delta - \delta_s \in \text{Der } U$. □

5.3 The abstract Jordan decomposition

Let $L \subset \mathfrak{gl}(V)$ be a semisimple linear Lie algebra. Theorem 5.4 says that L is isomorphic to $\text{Der } L \subset \mathfrak{gl}(L)$. We may thus view L as being embedded in both $\mathfrak{gl}(V)$ and $\mathfrak{gl}(L)$. In general, these enveloping algebras have different dimension. Our results up to this point give no guarantee that the semisimple and nilpotent parts $x_s, x_n \in \mathfrak{gl}(V)$ of any given $x \in L$ themselves lie in L, only that they exist. However, we do have this property for $L \cong \text{Der } L$ in $\mathfrak{gl}(L)$, by Lemma 5.5. Based on this, one might wonder if L does in fact have this property also in $\mathfrak{gl}(V)$. The next theorem says exactly this.

Theorem 5.6. *Let V be finite dimensional and let $L \subset \mathfrak{gl}(V)$ be a semisimple linear Lie algebra. Then L is closed under the Jordan decomposition in $\mathfrak{gl}(V)$, i.e. if $x \in L$ has Jordan decomposition $x = x_s + x_n$ in $\mathfrak{gl}(V)$, then $x_s, x_n \in L$.*

Proof. Let $x = x_s + x_n$ be as in the theorem, and write $N = N_{\mathfrak{gl}(V)}(L)$. We begin by expressing L as the intersection of certain subalgebras of $\mathfrak{gl}(V)$. First, V is an L-module under the action $(x, v) \mapsto x(v)$, $x \in L$, $v \in V$, with the corresponding representation being the inclusion map $L \to \mathfrak{gl}(V)$. Apply Weyl's Theorem on this representation to get V as direct sum of irreducible L-submodules, say $W_1, \ldots, W_n$. Now define, for $i = 1, \ldots, n$,

$$L_i = \{x \in \mathfrak{gl}(V) \mid x(W_i) \subset W_i,\ \text{tr}(x|_{W_i}) = 0\}.$$

We may verify that each of the two conditions $x(W_i) \subset W_i$, $\text{tr}(x|_{W_i}) = 0$ define a subalgebra of $\mathfrak{gl}(V)$ (to show closedness under the bracket for the latter, use the identities $\text{tr}(xy) = \text{tr}(yx)$ and $(xy)|_W = x|_W y|_W$); hence the L_i are subalgebras of $\mathfrak{gl}(V)$. Since each W_i is an L-submodule, $L(W_i) \subset W_i$, and L being semisimple implies $\text{tr}(L|_{W_i}) = \text{tr}([L, L]|_{W_i}) = \text{tr}([L|_{W_i}, L|_{W_i}]) = 0$. This shows that $L \subset L_i$, $i = 1, \ldots, n$. Next, take $L' = N_{\mathfrak{gl}(V)}(L) \cap (\bigcap_{i=1}^n L_i)$. Then L is an ideal of L' and L' is a subalgebra of N by construction. We now show that $L' = L$. First, L' is a (finite dimensional) L-module under the adjoint action $(x, y) \mapsto [xy]$, $x \in L$, $y \in L'$. Moreover, L is an L-submodule of L', so Weyl's Theorem furnishes an L-submodule M of L' such that $L' = L \oplus M$. This gives $L.L' = L.L \oplus L.M$, or equivalently $[LL'] = [LL] \oplus [LM]$. However, $[LL'] \subset L$ and $[LL] = L$, so this is only possible if $L.M = [LM] = 0$. The L-submodule

W_i has corresponding irreducible representation $|_{W_i} : L \to \mathfrak{gl}(W_i)$, and we have $[L|_{W_i}, M|_{W_i}] = [LM]|_{W_i} = 0$. Shur's Lemma now says that the algebra $M|_{W_i} \subset \mathfrak{gl}(W_i)$ consists of scalars. Furthermore, $M \subset L' \subset L_i$, so $\mathrm{tr}(M|_{W_i}) = 0$. Since $M|_{W_i}$ are scalars we must have $M|_{W_i} = 0$. But $V = W_i \oplus \cdots \oplus W_n$, so we must have $M = 0$ (since $M \subset \mathfrak{gl}(V)$), which means that $L' = L$.
Now let $x \in L$ have Jordan decomposition $x = x_s + x_n$ in $\mathfrak{gl}(V)$. From the above we know that $x \in L_i$ for all $i = 1, \ldots, n$. Then Theorem 3.2(d) gives $x_s(W_i) \subset W_i$ and similarly for x_n. Being nilpotent, x_n clearly satisfies $\mathrm{tr}(x_n|_{W_i}) = 0$, and then $0 = \mathrm{tr}(x|_{W_i}) = \mathrm{tr}(x_s|_{W_i})$. This says that $s_n, x_n \in L_i$ for all i. Moreover, the final remark of Section 3.1 says that since $x \in N_{\mathfrak{gl}(V)}(L)$ (which is equivalent to $\mathrm{ad}_{\mathfrak{gl}(V)}\ x(L) \subset L$) we have $\mathrm{ad}_{\mathfrak{gl}(V)}\ x_s(L) \subset L$ and similarly for x_n, i.e. $s_n, x_n \in N_{\mathfrak{gl}(V)}(L)$. This shows that $x_s, x_n \in L' = L$. □

As suggested in the previous section, Theorem 5.4 and Lemma 5.5 enables us to "carry over" the Jordan decomposition along ad_L. To be more precise,

Definition 5.2. Let L be semisimple. Given $x \in L$, there exists unique $s, n \in L$ such that $\mathrm{ad}_L\ x = \mathrm{ad}_L\ s + \mathrm{ad}_L\ n$ is the usual Jordan decomposition of $\mathrm{ad}_L\ x$ in $\mathfrak{gl}(L)$. We say that $x = s + n$ is the **abstract Jordan decomposition** of x.

Here $\mathrm{ad}_L\ s$ is semisimple and $\mathrm{ad}_L\ n$ nilpotent. By abuse of language, we call s the semisimple part of x, and n its nilpotent part. Of course, if L happens to be linear, then x has actual semisimple and nilpotent parts x_s, x_n. But $x_s, x_n \in L$ by Theorem 5.6, so we may send them across ad_L. By Lemma 3.1, $\mathrm{ad}_L\ x_s$ and $\mathrm{ad}_L\ x_n$ are semisimple and nilpotent, respectively, and $[\mathrm{ad}_L\ x_s, \mathrm{ad}_L\ x_n] = \mathrm{ad}_L\ [x_s x_n] = 0$. Thus, by the uniqueness of the Jordan decomposition (Theorem 3.2(b)), we must have $\mathrm{ad}_L\ x_s = \mathrm{ad}_L\ s$, $\mathrm{ad}_L\ x_n = \mathrm{ad}_L\ n$. In other words,

Corollary 5.6.1. *Let $L \subset \mathfrak{gl}(V)$ be a semisimple linear Lie algebra, and let $x \in L$. If $x = s + n$ and $x = x_s + x_n$ are, respectively, the abstract and usual Jordan decompositions of x, then $s = x_s$, $n = x_n$.*

We may therefore always use x_s and x_n to denote the semisimple and nilpotent parts of $x \in L$, with no distinction between what type of Jordan decomposition we are performing, since they always agree whenever both exist. Another important corollary to Theorem 5.6 is that this uniqueness, in a sense, carries over to homomorphic images of the abstract decomposition.

Corollary 5.6.2. *Let L be semisimple and $\phi : L \to \mathfrak{gl}(V)$ a finite dimensional representation of L. If $x \in L$ has (abstract) Jordan decomposition $x = x_s + x_n$, then $\phi(x) \in \mathfrak{gl}(V)$ has usual Jordan decomposition $\phi(x) = \phi(x_s) + \phi(x_n)$.*

Proof. Begin by writing $\mathcal{S} = \mathrm{ad}_{\phi(L)}\ \phi(x_s)$, $\mathcal{N} = \mathrm{ad}_{\phi(L)}\ \phi(x_n)$. Observe that $\mathrm{ad}_{\phi(L)}\ \phi(x) = \mathcal{S} + \mathcal{N}$ and $\mathcal{S}, \mathcal{N} \in \mathfrak{gl}(\phi(L))$. These commute:

$$[\mathcal{S}, \mathcal{N}] = \mathrm{ad}_{\phi(L)}\ [\phi(x_s), \phi(x_n)] = \mathrm{ad}_{\phi(L)}\ \phi([x_s x_n]) = 0.$$

Now x_n is nilpotent, i.e. $\mathrm{ad}_L\ x_n$ is. Then, for some positive integer m,

$$\mathcal{N}^m \phi(y) = (\mathrm{ad}_{\phi(L)}\ \phi(x_n))^m(\phi(y)) = \underbrace{[\phi(x_n)[\phi(x_n)\dots[\phi(x_n)}_{m}\phi(y)]]]$$
$$= \phi(\underbrace{[x_n[x_n\dots[x_n}_{m}y]]]) = \phi((\mathrm{ad}_L\ x_n)^m(y)) = 0, \quad (y \in L).$$

Hence $\mathcal{N}$ is nilpotent. Similarly, $\mathrm{ad}_L\ x_s$ is semisimple, i.e. it has an eigenvector basis $(x_i)_i$ of L. Then $(\phi(x_i))_i$ spans $\phi(L)$, and each $\phi(x_i)$ is an eigenvector of $\mathcal{S}$, which is easily verified. Therefore $\mathcal{S}$ is semisimple. Uniqueness now implies that $\mathrm{ad}_{\phi(L)}\ \phi(x) = \mathrm{ad}_{\phi(L)}\ \phi(x_s) + \mathrm{ad}_{\phi(L)}\ \phi(x_n)$ is the Jordan decomposition of $\mathrm{ad}_{\phi(L)}\ \phi(x)$ in $\mathfrak{gl}(\phi(L))$. Moreover, the algebra $\phi(L)$ is semisimple since it is the homomorphic image of a semisimple algebra. We may hence take the abstract Jordan decomposition of $\phi(x)$ in $\phi(L)$, which must be $\phi(x) = \phi(x_s) + \phi(x_n)$ by the above and Definition 5.2. Finally, Corollary 5.6.1 gives that this is the usual Jordan decomposition of $\phi(x)$ in $\mathfrak{gl}(V)$.

□

As an application, let L be semisimple and let K be a subalgebra of L. Consider the two representations $\mathrm{ad}_L : K \to \mathfrak{gl}(L)$ and $\mathrm{ad}_K : K \to \mathfrak{gl}(K)$. To say that $x \in K$ is semisimple in L is by definition to say that the operator $\mathrm{ad}_L\ x \in \mathfrak{gl}(L)$ is semisimple. The corollary implies that $\mathrm{ad}_K\ x \in \mathfrak{gl}(K)$—a different operator—is semisimple (note that Lemma 3.1(i) is not applicable since K need not be linear). We get a similar result when replacing "semisimple" with "nilpotent".

6 The Cartan decomposition

6.1 Toral subalgebras

We now introduce another family of Lie algebras. These are the *toral subalgebras*, and we have postponed their definition until now for the simple reason that they require the use of the abstract Jordan decomposition in their formulation.

Definition 6.1. Let L be semisimple. We say that a subalgebra T of L is a **toral subalgebra** if every element of T is semisimple.

If we on the opposite end of the spectrum have a subalgebra in which every element is nilpotent, then Engel's Theorem tells us that the subalgebra is nilpotent. Our first result on toral subalgebras provides a similar, though perhaps more unexpected insight.

Lemma 6.1. *If T is a toral subalgebra of L, then $[TT] = 0$; that is, T is abelian.*

Proof. Fix $x, y \in T$ and write $\mathcal{A} = \mathrm{ad}_T\ x$, $\mathcal{B} = \mathrm{ad}_T\ y$. Since x, y are semisimple in L by hypothesis, so are $\mathcal{A}, \mathcal{B}$ in $\mathfrak{gl}(T)$ as explained in the discussion following Corollary 5.6.2. Put $z = \mathcal{B}x$ and note that $z = -\mathcal{A}y$ since $[yx] = -[xy]$. Being semisimple, $\mathcal{A}, \mathcal{B}$ are diagonalizable in $\mathfrak{gl}(T)$, i.e. each have a set of eigenvectors that forms a base of T. Let $(x_i)_i$ be such a basis of eigenvectors of $\mathcal{B}$, with corresponding eigenvalues $(\lambda_i)_i$. Write $x = \sum_i a_i x_i$, $(a_i \in F)$, which gives $z = \mathcal{B}x = \sum_i a_i \lambda_i x_i$. Now suppose that z is nonzero. If y happens to be an eigenvector of $\mathcal{A}$ (with eigenvalue λ, say) then $\mathcal{B}z = -\mathcal{B}\mathcal{A}y = -\lambda\mathcal{B}y = -\lambda[yy] = 0$. In other words z (which is nonzero) is an eigenvector of $\mathcal{B}$ with eigenvalue 0, so $z = x_j$, $\lambda_j = 0$ for some j. However, this would imply that $x_i = \sum_{i \neq j} a_i \lambda_i x_i$, which contradicts the linear independence of $(x_i)_i$. Therefore y cannot be an eigenvector of $\mathcal{A}$ unless z is zero. This implies that $0 = -\mathcal{A}y$ whenever y is an eigenvector of $\mathcal{A}$, meaning y must have eigenvalue 0. This holds for any choice of $y \in T$ in the beginning of the proof, so every eigenvalue of $\mathcal{A}$ must be zero. Since $\mathcal{A}$ is diagonalizable, $\mathcal{A} = 0$. This holds for any choice of $x \in T$ in the beginning of the proof, so $[xy] = \mathcal{A}y = 0$ for all $x, y \in T$. □

Of special importance are the *maximal toral subalgebras*—the toral subalgebras not properly contained in any larger toral subalgebra. We use H to denote such algebras. In the next section, they will play a central role in understanding the structure of their enveloping algebra. Beyond being abelian, that is, each of its elements commutes with the entirety of H, it is also the case that *only* the elements of H have this latter property. In other words,

Proposition 6.2. *If H is a maximal toral subalgebra of L, then $C_L(H) = H$.*

In order to prove this proposition, we will need the following lemma.

Lemma 6.3. *If L is nilpotent, then every finite-dimensional nonzero ideal I intersects the center of L nontrivially, i.e. $I \cap Z(L) \neq 0$.*

Proof. Consider the adjoint action of L on itself. Being an ideal, I is an L-submodule of this action, whose corresponding representation $\phi : L \to \mathfrak{gl}(I)$ is given by $\phi(x) = (\mathrm{ad}_L\, x)|_I$, $(x \in L)$. As we have seen, L being nilpotent implies that every element of L is ad-nilpotent, so $\phi(L) \subset \mathfrak{gl}(I)$ is a linear Lie algebra of nilpotent operators, where I is finite-dimensional and nonzero. Theorem 8.1 thus furnishes a nonzero $x \in I$ such that $0 = \phi(L).x = [Lx]$, or equivalently, $x \in I \cap Z(L)$. □

We will also use the fact that the sum of any pair of commuting, semisimple linear operators is again semisimple (for a proof, see Theorem 8.8 of [6]), and the fact that the composition of two commuting nilpotent operators is nilpotent. Finally, we will use Corollary 6.4.1 of the next section, which of course will itself be proved without the help of Proposition 6.2, as to not result in circular reasoning.

Proof of Proposition 6.2. The inclusion $H \subset C_L(H)$ is clear from H being abelian. Let κ be the Killing form of L, and put $C = C_L(H)$. We split the proof into seven steps.

(1) C is closed under the abstract Jordan decomposition.

By definition, $x \in C$ if and only if $\mathrm{ad}_L\ x(H) = 0$. The semisimple and nilpotent parts x_s, x_n of x are by definition the unique elements of L such that $\mathrm{ad}_L\ x = \mathrm{ad}_L\ x_s + \mathrm{ad}_L\ x_n$ is the usual Jordan decomposition in $\mathfrak{gl}(L)$. But then $\mathrm{ad}_L\ x_s(H) = 0 = \mathrm{ad}_L\ x_n(H)$ by Theorem 3.2(d), so $x_s, x_n \in C$.

(2) If $x \in C$ is semisimple, then $x \in H$.

Write Fx for the span of x. We have $[Fx, Fx] = 0$, and also $[Fx, H] = 0$ by $x \in C$. Using this and Lemma 6.1, we see that $H + Fx$ is abelian (take its bracket with itself). The elements of $H + Fx$ are hence semisimple, since each is a sum of commuting semisimple operators. Then $H + Fx$ is toral, so maximality forces $H + Fx = H$, meaning $x \in H$.

(3) $\kappa|_{H\times H}$ is nondegenerate.

Suppose $h \in H$ is such that $\kappa(h, H) = 0$. Let $x \in C$. By (1), $x_n \in C$, so $\mathrm{ad}_L\ h$, $\mathrm{ad}_L\ x_n$ commute. Then, since $\mathrm{ad}_L\ x_n$ is nilpotent, $\mathrm{ad}_L\ h\ \mathrm{ad}_L\ x_n$ is nilpotent. This shows that $\kappa(h, x_n) = \mathrm{tr}(\mathrm{ad}_L\ h\ \mathrm{ad}_L\ x_n) = 0$, or equivalently $\kappa(h, x) = \kappa(h, x_s)$. But $x_s \in H$ by (2), so we see that $\kappa(h, x) = 0$. Since x was arbitrary, $\kappa(h, C) = 0$. This forces $h = 0$ since $\kappa|_{C\times C}$ is nondegenerate (Corollary 6.4.1), which proves the statement.

(4) C is nilpotent.

We use Engel's Theorem. Fix $x \in C$; since $x_s, x_n \in C$ by (1), we may write $\mathrm{ad}_C\ x = \mathrm{ad}_C\ x_s + \mathrm{ad}_C\ x_n$. But (2) says that $x_s \in H$, which implies $\mathrm{ad}_C\ x(C) = [xC] \subset [HC] = 0$, or in other words $\mathrm{ad}_C\ x_s = 0$. Therefore $\mathrm{ad}_C\ x = \mathrm{ad}_C\ x_n$ is nilpotent (Corollary 5.6.2) and we are done.

(5) $H \cap [CC] = 0$

We have $\kappa(H, [CC]) = \kappa([HC], C) = \kappa(0, C) = 0$. Part (3) says that no nonzero $h \in H$ may satisfy $\kappa(H, h) = 0$—hence, $H \cap [CC] = 0$.

(6) C is abelian.

Suppose not, i.e. $[CC]$ is nonzero. It is an ideal of C and the latter is nilpotent by (4), so there exists a nonzero $x \in [CC] \cap Z(C)$ by Lemma 6.3. If this x equals its semisimple part, then $x \in H$ by (2), and then $x = 0$ by (5). This is not possible, so we must have $x_n \neq 0$. $x \in Z(C)$ is equivalent to $\mathrm{ad}_L\ x(C) = 0$ and as in part (1) we thus obtain $\mathrm{ad}_L\ x_n(C) = 0$, meaning $x_n \in Z(C)$. Thus $\mathrm{ad}_L\ x_n, \mathrm{ad}_L\ C$ commute, and arguing as in part (3) lets us conclude that $\kappa(x_s, C) = 0$. But $x_n \in C$ by (1) and is nonzero, which is a contradiction since $\kappa|_{C\times C}$ is nondegenerate.

(7) $C = H$.

Suppose not, so there exists an $x \in C \setminus H$. If it equals its semisimple part, then $x \in H$ by (2)— a contradiction—so $x_n \neq 0$. By (1), $x_n \in C$. Then $\kappa_C(x_n, C) = \mathrm{tr}(\mathrm{ad}_C\ x_n\ \mathrm{ad}_C\ C) = 0$ since $\mathrm{ad}_C\ x_n$ is nilpotent and, by (6), commutes with $\mathrm{ad}_C\ C$. We now have the same contradiction as for (6).

□

In the next section, we put maximal toral subalgebras to use in understanding the structure of its enveloping semisimple algebra, in a way that is reminiscent of how one decomposes a vector space into the eigenspaces of a given semisimple (diagonalizable) operator.

6.2 Roots of a semisimple algebra

Let L be a nonzero Lie algebra and let K be a subalgebra of L. Suppose $x \in L$, $x \neq 0$ is a simultaneous eigenvector of $\mathrm{ad}_L\ K$ in the sense that there exist associated scalars $\lambda_{x,h} \in F$, $h \in K$ such that $\mathrm{ad}_L\ h(x) = \lambda_{x,h}x$ for all $h \in K$. Since the $\lambda_{x,h}$ are eigenvalues and therefore unique, any simultaneous eigenvector x defines a unique map $\alpha_x : K \to F$ given by $\alpha_x(h) = \lambda_{x,h}$. The linearity of ad_L implies

$$\mathrm{ad}_L\ (ah + bk)(x) = a\alpha_x(h) + b\alpha_x(k), \quad h, k \in K;\ a, b \in F,$$

so α_x is in fact a linear functional, i.e. $\alpha_x \in K^*$. In other words, given x as above, the problem of determining $\alpha \in K^*$ such that

$$[hx] = \alpha(h)x \quad \text{for all } h \in K$$

has a unique solution $\alpha_x \in K^*$. Of course, if $y \in L$, $\alpha \in K^*$ satisfies the above condition, then y is a common eigenvector as defined earlier, and $\alpha = \alpha_y$. Motivated by this, we define for any $\alpha \in K^*$ the set

$$L_\alpha = \{x \in L \mid [hx] = \alpha(h)x \text{ for all } h \in K\}.$$

Note that $0 \in L_\alpha$ for all $\alpha \in K^*$. It is straightforward to verify that L_α is a subspace of L. By the above, $x \in L_\alpha$, $x \neq 0$ if and only if x is a simultaneous eigenvector of $\mathrm{ad}_L\ K$, and $\alpha = \alpha_x$. From this we also see that the subspaces are disjoint: If $x \in L_\alpha$, $y \in L_{\alpha'}$, $\alpha \neq \alpha'$ then $\alpha_x = \alpha \neq \alpha' = \alpha_y$, so $x \neq y$. Moreover, if L is finite dimensional, then the set

$$\Phi = \{\alpha \in K^* \mid \alpha \neq 0,\ L_\alpha \neq 0\}$$

is finite (otherwise L would contain an infinite number of disjoint subspaces). In this case we call Φ the set of *roots* of L relative to K.

Now suppose $L \neq 0$ is semisimple. If every element of L were nilpotent, then every element would be ad-nilpotent (Corollary []), and then Engel's Theorem would imply that L is nilpotent. But $[LL] = L$, so this cannot be. Hence L contains elements with nonzero semisimple part. Take the span of any such semisimple part and the result is a toral subalgebra, so L *must* have maximal toral subalgebras. Fix H to be one such subalgebra. By Lemma 6.1, $\mathrm{ad}_L\ H$ is a set of commuting linear operators of a finite dimensional vector space, and therefore simultaneously diagonalizable (for a proof, see Theorem 2 of [2]). This means that L has a base of simultaneous eigenvectors, say $(x_i)_i$. But $x_i \in L_{\alpha_{x_i}}$, so we have the direct sum

$$L = L_0 \oplus \bigoplus_{\alpha \in \Phi} L_\alpha. \tag{6.1}$$

This is the *Cartan decomposition* of L with respect to H.

Proposition 6.4. *Let L be semisimple and let $\alpha, \beta \in H^*$. Then*

(i) $[L_\alpha L_\beta] \subset L_{\alpha+\beta}$*;*

(ii) if $\alpha \neq 0$ and $x \in L_\alpha$, then x is ad-nilpotent;

(iii) if $\alpha + \beta \neq 0$, then $\kappa(L_\alpha, L_\beta) = 0$.

Proof. Let $x \in L_\alpha$, $y \in L_\beta$, $h \in H$. By the Jacobi identity,

$$[h[xy]] = [[hx]y] + [x[hy]] = \alpha(h)[xy] + \beta(h)[xy] = (\alpha+\beta)(h)[xy],$$

which proves (i). Using (i) gives $(\mathrm{ad}_L\ x)^n(L_\beta) \subset L_{n\alpha+\beta} = 0$ for sufficiently large n (since Φ is finite), and then $(\mathrm{ad}_L\ x)^n(L) = 0$ for sufficiently large n by (6.1), so (ii) is true. To show (iii), take $h \in H$ such that $(\alpha+\beta)(h) \neq 0$, which is possible by hypothesis. By associativity,

$$\begin{aligned}(\alpha+\beta)(h)\kappa(L_\alpha, L_\beta) &= \kappa(\alpha(h)L_\alpha, L_\beta) + \kappa(L_\alpha, \beta(h)L_\beta) \\ &= \kappa([hL_\alpha], L_\beta) + \kappa(L_\alpha, [hL_\beta]) \\ &= -\kappa([L_\alpha h], L_\beta) + \kappa([L_\alpha h], L_\beta) = 0,\end{aligned}$$

and it follows that $\kappa(L_\alpha, L_\beta)$ must be zero. □

Observe that $L_0 = \{x \in L \mid [hx] = 0 \text{ for all } h \in H\} = C_L(H)$. In the previous section we used the fact that the restriction of the Killing form to $C_L(H)$ is nondegenerate, which we prove now.

Corollary 6.4.1. *Let H be a maximal toral subalgebra of L. Write $C = C_L(H)$. The restriction $\kappa|_{C\times C}$ is nondegenerate.*

Proof. Suppose $x \in L_0$, $\kappa(x, L_0) = 0$. From (iii) we see that $\kappa(x, L_\beta) = 0$ for all $\beta \in \Phi$ (recall that $\beta \neq 0$ by definition). Hence x is orthogonal to every summand in (6.1), i.e. $\kappa(x, L) = 0$, so κ being nondegenerate forces $x = 0$. □

We are now positioned to use Proposition 6.2 and conclude that

Corollary 6.4.2. *The restriction $\kappa|_{H\times H}$ is nondegenerate.*

In view of the Cartan decomposition, information about the roots could potentially be translated into information about the structure of L, so it is natural to study these further. In doing this, we start by "embedding" Φ in H, in an appropriate fashion. Recall that a nondegenerate bilinear form β on a finite-dimensional vector space V furnishes a natural isomorphism $V^* \cong V$, which associates to each linear functional $f \in V^*$ a unique vector $v_f \in L$ satisfying $\beta(v_f, v) = f(v)$ for all $v \in V$. Since the Killing form is nondegenerate on H, we hence associate to each $\phi \in H^*$ a unique element $t_\phi \in H$ satisfying $\kappa(t_\phi, h) = \phi(h)$ for all $h \in H$. Conversely, every $h \in H$ has a $\phi_h \in H^*$ such that $t_{\phi_h} = h$. Under this identification, we can prove that

Proposition 6.5. *Φ spans H^*.*

Proof. By the above, Φ spans H^* if and only if $H_\Phi = \{t_\alpha \mid \alpha \in \Phi\}$ spans H. Suppose not, so that the set $(H_\Phi)^\perp = \{h \in H \mid \kappa(h, H_\Phi) = 0\}$ is nonempty. The definition of t_α implies that $h \in (H_\Phi)^\perp$ if and only if $\kappa(h, t_\alpha) = \kappa(t_\alpha, h) = \alpha(h) = 0$ for all $\alpha \in \Phi$. If so, then $[hx] = \alpha(h) = 0$ for all $x \in L_\alpha$, $\alpha \in \Phi$, or $[xL_\alpha] = 0$ for all $\alpha \in \Phi$. But $[hH] = 0$ (H is abelian), so then $[hL] = 0$ by (6.1), and therefore $h \in Z(L) = 0$. This is, of course, contrary to our assumption. □

As an immediate application, the proposition enables us to select a set of roots $\{\alpha_1, \ldots, \alpha_\ell\}$ as basis of H^* (i.e. $\alpha_i \in \Phi$ for all $1 \leq i \leq \ell$). We say that the number $\ell = \dim H^* = \dim H$ is the *rank* of Φ. By (6.1),

$$\dim L = \ell + \sum_{\alpha \in \Phi} \dim L_\alpha, \tag{6.2}$$

so if we can compute the dimensions of the spaces L_α, $\alpha \in \Phi$ along with the size of Φ, we obtain admissible values of $\dim L$ (as a step in this direction, (a) of the next proposition says that $|\Phi|$ must be even).

6.3 Root spaces, actions, and weights

In proceeding with our investigation, we would like to employ the adjoint representation, as is our usual modus operandi. The action of L on itself is to some extent described by Proposition 6.4(i), but this description is rather coarse. Luckily, we can say more in the case where a space L_α, ($\alpha \in \Phi$) is acting adjointly on the related space $L_{-\alpha}$, which is a good starting point. Let us call L_α a *root space* if it is nonzero, or equivalently, if α is a root.

Proposition 6.6. *Let $\alpha \in \Phi$, and define $H_\alpha = [L_\alpha L_{-\alpha}]$. Then*

(a) $-\alpha \in \Phi$, i.e. $L_{-\alpha}$ is a root space;

(b) $[xy] = \kappa(x, y)t_\alpha$ for all $x \in L_\alpha$, $y \in L_{-\alpha}$;

(c) $H_\alpha = Ft_\alpha$ (in particular, $\dim H_\alpha = 1$).

Proof. (a) Proposition 6.4(iii) says that $\kappa(L_\alpha, L_\beta) = 0$ for all $\beta \in H^* \setminus \{-\alpha\}$. Hence, if $-\alpha \notin \Phi$, or equivalently $L_{-\alpha} = 0$, then $\kappa(L_\alpha, L_{-\alpha}) = 0$, which in turn gives $\kappa(L_\alpha, L) = 0$ by (6.1). This contradicts the nondegeneracy of κ.

(b) Write $k = [xy] - \kappa(x, y)t_\alpha$; we have $k \in H$ since $[xy] \in [L_\alpha L_{-\alpha}] \subset L_0 = H$ and $t_\alpha \in H$. For all $h \in H$,

$$\kappa(h, [xy]) = \kappa([hx], y) = \alpha(h)\kappa(x, y) = \kappa(t_\alpha, h)\kappa(x, y) = \kappa(h, \kappa(x, y)t_\alpha).$$

In other words $\kappa(H, k) = 0$, and then $k = 0$ by nondegeneracy.

(c) Take any nonzero $x \in L_\alpha$. Having $\kappa(x, L_{-\alpha}) = 0$ would as in the proof of (a) contradict the nondegeneracy of κ, so there exist nonzero $y \in L_{-\alpha}$ such that $\kappa(x, y) \neq 0$. Then $[L_\alpha L_{-\alpha}] \neq 0$ by (b), which, again by (b), forces $H_\alpha := [L_\alpha L_{-\alpha}] = F t_\alpha$.

□

Remark. Part (a) allows us to (arbitrarily) choose one root out of each pair $\pm\alpha \in \Phi$. Gathering them in a set Φ', we have $\Phi \setminus \Phi' = -\Phi'$ and thus $|\Phi| = 2|\Phi'|$. We may also without loss of generality assume that the basis chosen in the previous section satisfies $\{\alpha_1, \ldots, \alpha_\ell\} \subset \Phi'$.

With this proposition in hand, we take our identification of H and its dual a step further; we define a nondegenerate symmetric bilinear form $(-,-)$ on H^* by letting $(\gamma, \delta) = \kappa(t_\gamma, t_\delta)$ for all $\gamma, \delta \in H^*$. The reason we postponed doing so until now is that the proposition enables us to prove that

Lemma 6.7. $(\alpha, \alpha) \neq 0$ *for all* $\alpha \in \Phi$.

Proof. Suppose not, so $(\alpha, \alpha) = \kappa(t_\alpha, t_\alpha) = \alpha(t_\alpha) = 0$ for some $\alpha \in \Phi$. Find x, y as in the proof of (c) and, if needed, interchange one of them with a scalar multiple such that $\kappa(x, y) = 1$. Thus, $[xy] = t_\alpha$. Let $S = \operatorname{span}_F\{t_\alpha, x, y\}$, which is three-dimensional since t_α, x, y lie in different root spaces. By assumption, $[t_\alpha x] = \alpha(t_\alpha)x = 0$ and similarly for y. It follows that $S^{(1)} = F t_\alpha$ and $S^{(2)} = 0$, so S a solvable subalgebra of L. Hence, $[SS] = F t_\alpha$ is nilpotent (Corollary 4.8.1), which is equivalent to $\operatorname{ad}_L t_\alpha$ being nilpotent. But we know $\operatorname{ad}_L t_\alpha$ to be semisimple (since $t_\alpha \in H$), which forces $\operatorname{ad}_L t_\alpha = 0$, or equivalently, $t_\alpha \in Z(L) = 0$. This is impossible since all t_α, $\alpha \in \Phi$ are nonzero by construction. □

Interestingly, in proving the above, we procured a subalgebra S of L that contains elements of both the maximal toral subalgebra and two different root spaces, and whose bracket multiplication table we now know to be

$$[t_\alpha x] = (\alpha, \alpha)x \neq 0, \quad [t_\alpha y] = -(\alpha, \alpha)y \neq 0, \quad [xy] = t_\alpha.$$

We may of course scale any of the three basis elements to our liking. Clearly, the next logical step would be to let S act on L in some appropriate fashion, since we by this point have plenty of knowledge on how each of the elements used in defining S behave when bracketed with arbitrary elements of L. Before we do this, however, we sharpen our understanding of S with another proposition.

Proposition 6.8. *Let* $x_\alpha \in L_\alpha$ *be a nonzero element of a root space. There exists* $y_\alpha \in L_{-\alpha}$ *such that, after setting* $h_\alpha := [x_\alpha y_\alpha]$*, we have*

$$S_\alpha := \operatorname{span}_F\{h_\alpha, x_\alpha, y_\alpha\} \cong \mathfrak{sl}(2, F).$$

This isomorphism is given by the mapping $h_\alpha \mapsto h$, $x_\alpha \mapsto x$, $y_\alpha \mapsto y$, *where the latter vectors are the standard basis of* $\mathfrak{sl}(2, F)$. *Moreover,* $h_\alpha = \frac{2}{(\alpha,\alpha)} t_\alpha = -h_{-\alpha}$ *(in particular,* $H_\alpha = F h_\alpha$*).*

If the notation $\mathfrak{sl}(2, F)$ is unfamiliar, we direct the reader to the appendix, and more specifically, Section 8.2.

Proof. Let nonzero $x_\alpha \in L_\alpha$ be given. Again find $y_\alpha \in L_\alpha$ as in the proof of (c) and, if necessary, rescale them in such a way that $\kappa(x_\alpha, y_\alpha) = \frac{2}{(\alpha,\alpha)}$ (this is possible in view of Lemma 6.7). For this choice, $h_\alpha = \frac{2}{(\alpha,\alpha)} t_\alpha$ by (b), and substituting h_α for t_α in the preceding multiplication table immediately yields

$$[h_\alpha x_\alpha] = 2x_\alpha, \quad [h_\alpha y_\alpha] = -2y_\alpha, \quad ([x_\alpha y_\alpha] =: h_\alpha). \tag{6.3}$$

Since these equations characterize a bracket identical to the one of $\mathfrak{sl}(2, F)$, the two algebras are necessarily isomorphic under the prescribed mapping. Finally,

$$h_{-\alpha} = \frac{2}{(-\alpha, -\alpha)} t_{-\alpha} = \frac{2}{(\alpha, \alpha)} t_{-\alpha} = -\frac{2}{(\alpha, \alpha)} t_\alpha = -h_\alpha,$$

where the third equality follows from the fact that $\kappa(t_\alpha + t_{-\alpha}, h) = \kappa(t_\alpha, h) + \kappa(t_{-\alpha}, h) = \alpha(h) - \alpha(h) = 0$ for all $h \in H$, or in other words, $t_\alpha + t_{-\alpha} = 0$. □

With this we fix $\alpha \in \Phi$, construct S_α, and let S_α act on L through the adjoint action. We will study how S_α acts on several S_α-submodules of L, starting with L itself. Let x be an eigenvector of $\mathrm{ad}_L\, h_\alpha$, with eigenvalue λ. Using the Cartan decomposition, we can write x uniquely of the form

$$x = h + \sum_{\beta \in \Phi} x_\beta, \quad (h \in H,\ x_\beta \in L_\beta). \tag{6.4}$$

Since x is nonzero, either x lies in H or at least one of the x_β is nonzero. In the former case, $\mathrm{ad}_L\, h_\alpha(x) = 0$ by Lemma 6.1, meaning $\lambda = 0$ (and since H indeed is nonzero, we see that 0 always is an eigenvalue of $\mathrm{ad}_L\, h_\alpha$). Now suppose the latter—say, $x_\gamma \neq 0$, $\gamma \in \Phi$. Applying $\mathrm{ad}_L\, h_\alpha$ to both sides of (6.4) yields

$$\lambda h + \sum_{\beta \in \Phi} \lambda x_\beta = \lambda x = \mathrm{ad}_L\, h_\alpha(x) = 0 + \sum_{\beta \in \Phi} [h_\alpha x_\beta] = \sum_{\beta \in \Phi} \beta(h_\alpha) x_\beta,$$

and we deduce by uniqueness of this representation that $\lambda x_\beta = \beta(h_\alpha) x_\beta$ for all $\beta \in \Phi$. Hence, $\lambda = \gamma(h_\alpha)$. Conversely, any nonzero $x \in L_\beta$, $\beta \in \Phi$ is an eigenvector of $\mathrm{ad}_L\, h_\alpha$ with eigenvalue $\beta(h_\alpha)$, as is easy to verify. We therefore see that the eigenvalues of $\mathrm{ad}_L\, h_\alpha$ are exactly the scalars described by the set

$$\{0\} \cup \{\beta(h_\alpha) \mid \beta \in \Phi\}. \tag{6.5}$$

In view of Proposition 6.8,

$$\beta(h_\alpha) = \frac{2}{(\alpha, \alpha)} \beta(t_\alpha) = \frac{2}{(\alpha, \alpha)} \kappa(t_\beta, t_\alpha) = 2\frac{(\beta, \alpha)}{(\alpha, \alpha)} = -\beta(h_{-\alpha}).$$

For convenience, write $\langle \beta, \alpha \rangle = 2(\beta, \alpha)/(\alpha, \alpha)$ for all $\alpha, \beta \in \Phi$ (note that $\langle -, - \rangle$ is linear only in its first argument). With this, (6.5) becomes

$$\{0\} \cup \{\pm\langle\beta,\alpha\rangle \mid \beta \in \Phi'\}. \tag{6.6}$$

In the terminology of Section 8.2, the set (6.6) are exactly the so-called weights of h_α in L. Since S_α is isomorphic to $\mathfrak{sl}(2,F)$, *Corollary 8.6.1(a) ensures that these are integers.* We call the numbers $\langle\beta,\alpha\rangle$, $(\alpha,\beta \in \Phi)$ the *Cartan integers* of the decomposition.

Now construct an S_α-submodule of L, as follows. Given any $c \in F^*$, we note that the composition $c\alpha$ makes sense and is a linear functional, so we may consider the family of spaces $L_{c\alpha}$, $c \in F^*$. Define M by

$$M = \operatorname{span}_F \bigcup_{c \in F^*} L_{c\alpha}.$$

To verify that M is an S_α-submodule, first use Proposition 6.4(i) to see that

$$\begin{aligned}
&h_\alpha.L_{c\alpha} \subset [L_0 L_{c\alpha}] \subset L_{c\alpha} \subset M,\\
&x_\alpha.L_{c\alpha} \subset [L_\alpha L_{c\alpha}] \subset L_{(c+1)\alpha} \subset M,\\
&y_\alpha.L_{c\alpha} \subset [L_{-\alpha} L_{c\alpha}] \subset L_{(c-1)\alpha} \subset M, \quad (c \in F^*).
\end{aligned}$$

where $1 \in F^*$ is the identity functional. Thus, $S_\alpha.L_{c\alpha} \subset M$ for all $c \in F^*$. An arbitrary element of M is a finite linear combination in which the i:th summand is an element of some $L_{c_i\alpha}$, $c_i \in F^*$, so it follows from the preceding that $S_\alpha.M \subset M$, and the claim follows. Of course, only a finite number of spaces in the family $L_{c\alpha}$, $c \in F^*$ are nonzero, namely H together with all root spaces, i.e. the spaces for which $c\alpha \in \Phi$. We may therefore write M as a direct sum of nonzero spaces

$$M = H \oplus L_{c_1\alpha} \oplus \cdots \oplus L_{c_t\alpha} \tag{6.7}$$

for some particular choice of nonzero distinct functionals $c_1, \ldots, c_t \in F^*$. We now turn our attention to the weights of h_α in M. These are just the eigenvalues of $\mathrm{ad}_L\ h_\alpha$, but under the additional constraint that the corresponding eigenvectors are taken from M. Evidently, these weights are then a subset of (6.6), meaning each is 0 and/or a Cartan integer. Now let $x \in M$ be arbitrary, and use (6.7) to write x uniquely of the form $x = h + x_1 + \cdots + x_t$, $(h \in H, x_i \in L_{c_i\alpha})$. Proposition 6.8 says that $h_\alpha = 2t_\alpha/(\alpha,\alpha)$, which allows us to compute that

$$\mathrm{ad}_L\ h_\alpha(x_i) = (c_i\alpha)(h_\alpha)x_i = c_i\alpha(h_\alpha)x_i = c_i\frac{2\alpha(h_\alpha)}{(\alpha,\alpha)}x_i = c_i 2x = 2c_i x. \tag{6.8}$$

(In the final equality, we evaluated $c_i \in F^*$ at $2 \in F$. For convenience, we will write c_i for both the linear functional and its corresponding scalar when evaluated at $1 \in F$.) If we apply (6.8) under the assumption that x is an eigenvector of $\mathrm{ad}_L\ h_\alpha$ with eigenvalue λ, we obtain the equality

$$\lambda h + \lambda x_1 + \cdots + \lambda x_t = \lambda x = \mathrm{ad}_L\ h_\alpha(x) = 0 + 2c_1x_1 + \cdots + 2c_tx_t. \tag{6.9}$$

Both sides are unique representations of the same element of M, so $\lambda h = 0$ and $\lambda x_i = 2c_ix_i$ for all $1 \le i \le t$. We see that $h \neq 0$ implies $\lambda = 0$ and that $x_k \neq 0$,

$1 \leq k \leq t$ implies $\lambda = 2c_k$. But the scalars $c_1, \ldots, c_t$ are distinct and nonzero, which clearly means that most one "component" of x can be nonzero, in which case λ is the coefficient of this component in the right-hand side of (6.9). This is indeed the case since x is nonzero. Conversely, we may pick any nonzero x from one of the summands of (6.7) and obtain such an eigenvector. In summary,

Lemma 6.9. *The weights of h_α in M are integers $0, 2c_1, \ldots, 2c_t$, ($c_i \in F$ distinct and nonzero), and M is the direct sum of weight spaces*

$$M = M_0 \oplus M_{2c_1} \oplus \cdots \oplus M_{2c_t},$$

where $M_0 = H$ and $M_{2c_i} = L_{c_i\alpha}$.

Finally, we let S_α act on itself. It is by construction contained in the span of the three spaces $H, L_\alpha, L_{-\alpha}$, and these are in turn contained in (6.7), so S_α *is an S_α-submodule of M*. Similarly to before, all weights of h_α in S_α are a subset of the weights described by the above lemma. In fact, we can explicitly compute these; they are exactly $-2, 0, 2$, as can be deduced from the equations (6.3). Of course, the weights -2 and 2 correspond, respectively, to $c_{i_1} = -1$ and $c_{i_2} = 1$ for some $1 \leq i_1, i_2 \leq t$. In other words, they correspond to the root spaces $L_{-\alpha}$ and L_α, which we already know are summands of (6.7). It remains to be seen if there are any other possibilities for the c_i. This is solved by our next proposition, which also makes more precise some of our earlier results.

Proposition 6.10. *For all $\alpha \in \Phi$,*

(a) $\dim L_\alpha = 1$ *(in particular, $S_\alpha = H_\alpha \oplus L_\alpha \oplus L_{-\alpha}$);*

(b) $-\alpha \in \Phi$*, but no additional multiple of α is a root;*

(c) given any nonzero $x_\alpha \in L_\alpha$, there exists a unique $y_\alpha \in L_{-\alpha}$ such that $h_\alpha := \frac{2}{(\alpha,\alpha)} t_\alpha = [x_\alpha y_\alpha]$.

Proof. (a) Since the action is the adjoint action, S_α being simple implies that it is an irreducible S_α-submodule of M. If we use Weyl's Theorem to decompose M into irreducible S_α-submodules, we may thus without loss of generality assume that S_α is included as a summand. Keeping this in mind, we try to find more irreducible submodules. We computed that $\alpha(h_\alpha) = 2$ in one of the steps of (6.9), and if we only care that this value is nonzero, we can just write $Fh_\alpha \cap \operatorname{Ker} \alpha = 0$. It turns out that the dimensions of these two spaces add up to $\dim H$, as evidenced by the nondegeneracy of $\kappa|_H$ and the fact that

$$\begin{aligned}
(Fh_\alpha)^\perp &= \{h \in H \mid \kappa(h, Fh_\alpha) = 0\} \\
&= \{h \in H \mid \kappa(h, t_\alpha) = 0\} \\
&= \{h \in H \mid \kappa(t_\alpha, h) = 0\} \\
&= \{h \in H \mid \alpha(h) = 0\} \\
&= \operatorname{Ker} \alpha.
\end{aligned}$$

Hence, $H = Fh_\alpha \oplus \text{Ker }\alpha$. It it easy to verify that $S_\alpha.\text{Ker }\alpha = 0$, which means that Ker α may be decomposed into $\dim \text{Ker }\alpha = \dim H - 1 = \ell - 1$ one-dimensional irreducible S_α-submodules of M (simply choose a basis of Ker α and take each submodule to be the span of one of the basis elements). In other words, M can without loss of generality be decomposed as

$$M = S_\alpha \oplus W_1 \oplus \cdots \oplus W_{\ell-1} \oplus W, \tag{6.10}$$

where the first ℓ summands are irreducible—note also that they contain $H = M_0$—and W decomposes further as direct sum of irreducible submodules.

Now, recall that the weights of h_α in M are integers. Theorem 8.6 says that if any of the irreducible submodules of W has an even weight, then the same irreducible submodule has 0 as one of its weights, and should furnish a nonzero element of M_0, but this is clearly impossible. More generally, for M to have even weight m, it must have an eigenvector x of eigenvalue m, which we can write uniquely as a sum of elements x_i, where each x_i hails from one of the irreducible submodules of (6.10). Arguing as we did after our derivation of (6.8), one can then show that each element x_i must be an eigenvector with even eigenvalue m, or else be zero. Since we have excluded W from having such eigenvectors, and each W_i only has eigenvectors of eigenvalue 0, we see that if m is nonzero, then necessarily $x \in S_\alpha$. Hence, we conclude that *the only even weights of* h_α *in* M *are* $-2, 0, 2$.

According to Lemma 6.9, $2\alpha \in \Phi$ if and only if $c_i = 2$ for some $1 \le i \le t$. This would mean that $2c_i = 4$ is a weight, but we know from the preceding that this is not the case. In other words, $(2\Phi) \cap \Phi = \emptyset$. This also excludes $\frac{1}{2}\alpha$ from being a root, since $2(\frac{1}{2}\alpha) = \alpha \in \Phi$ by assumption. In the language of the lemma, we would say that $\frac{1}{2}$ is not one of the c_i. *Therefore* 1 *is not a weight*, i.e. $M_1 = 0$.

Corollary 8.6.1(c) lists a formula for computing the number of irreducible summands of (6.10). Using it now, we obtain that this number is

$$\dim M_0 + \dim M_1 = \dim M_0 = \dim H = \ell$$

which tells us that $W = 0$ and $M = S_\alpha \oplus \text{Ker }\alpha = \text{span}_F\{x_\alpha, y_\alpha\} \oplus H$. But at the same time, M should decompose into a direct sum of H and root spaces, as described by Lemma 6.9. The only way to reconcile these viewpoints is if $L_\alpha = Fx_\alpha$ and $L_{-\alpha} = Fy_\alpha$. In particular, $\dim L_\alpha = 1$.

(b) This is immediate from the definition of M and the proof of (a), in which we found that $M = H \oplus L_\alpha \oplus L_{-\alpha}$.

(c) It follows from $S_\alpha = [L_\alpha L_{-\alpha}] \oplus L_\alpha \oplus L_{-\alpha}$ and the one-dimensionality of the summands that the the way we chose y_α in the proof of Proposition 6.8 was the only way we could have done so.

□

Part (a) is perhaps the most attention-capturing, since it allows us to sharpen our estimation (6.2) into

$$\dim L = \ell + 2|\Phi'| \geq \ell + 2\ell = 3\ell.$$

With respect to dimension, a semisimple Lie algebra is therefore at least three times larger than its greatest maximal toral subalgebra(s). An example of a semisimple algebra that attains this lower bound is the familiar algebra $\mathfrak{sl}(2, F)$ along with its unique maximal toral subalgebra $H = Fh$ (one technically has to prove the last part, but we allow ourselves to assume so here for the sake of giving an example). Speaking of $\mathfrak{sl}(2, F)$, part (a) also says that we with ease can construct copies of it by forming the direct sum $[L_\alpha L_{-\alpha}] \oplus L_\alpha \oplus L_{-\alpha}$ for any root $\alpha \in \Phi$. A consequence of this is that any three-dimensional semisimple Lie algebra L is isomorphic to $\mathfrak{sl}(2, F)$: L is not abelian (recall that $[LL] = L$), so its maximal toral subalgebras must be proper (this of course holds in general). Choose one of these to be H. By the Cartan decomposition and H being proper, L must have at least one root space. We may then construct a copy of $\mathfrak{sl}(2, F)$, which necessarily must equal L since their dimensions agree.

6.4 Root systems

While we have made important progress in understanding the Cartan decomposition and root spaces, there is one central aspect of roots that we have yet to investigate. This aspect is how a given root relates to any root other than its negative. Although a complete classification of the admissible configurations of roots will elude us, we remark that such a classification not only exists, but extends to and allows for a complete classification of semisimple Lie algebras. We will in this section derive the final results necessary for this classification to be tractable. Such a classification is possible because, informally speaking, *roots may as well just be considered as sets of vectors of some euclidean space.* We make this more precise later; for now, we uncover some of the algebraic structure of Φ by studying sums of roots.

Suppose we have a triplet (L, H, Φ) as in the previous section and that we have fixed a basis $\{\alpha_1, \ldots, \alpha_\ell\} \subset \Phi$ of H^*, where ℓ by definition is the rank of Φ. Let $\beta \in \Phi$ be an arbitrary root and write it in this basis, say,

$$\beta = \sum_{k=1}^{\ell} c_k \alpha_k, \quad c_k \in F. \tag{6.11}$$

Recall that each $\langle -, \alpha_i \rangle$, $(1 \leq i \leq \ell)$ is a linear functional on H that moreover is integer-valued on Φ. Applying these linear functionals to both sides of (6.11) yields the ℓ equations

$$\langle \beta, \alpha_i \rangle = \sum_{k=1}^{\ell} \langle \alpha_k, \alpha_i \rangle c_k, \quad 1 \leq i \leq \ell$$

which we can express simultaneously as a matrix equation

$$b = Ac$$

where $b = (\langle \beta, \alpha_i \rangle)_i$, $A = (\langle \alpha_j, \alpha_i \rangle)_{ij}$, $c = (c_i)_i$, and the first two of these have integral entries. Suppose A is *not* invertible, which is equivalent to saying that there exists a nonzero $x \in F^\ell$ such that $Ax = 0$. We can exploit this assumption to define a nonzero linear functional $\phi = \sum_{k=1}^{\ell} x_k \alpha_k$ by letting its coefficients be given by $(x_i)_i = x$. This linear functional satisfies

$$0 = Ax = (\langle \alpha_i, \alpha_j \rangle)_{ij} \times (x_i)_i = \left(\sum_{k=1}^{\ell} \langle \alpha_k, \alpha_i \rangle x_k \right)_i = (\langle \phi, \alpha_i \rangle)_i.$$

This is the case if and only if $(\phi, \alpha_i) = 0$ for all $1 \leq i \leq \ell$ by the definition of $\langle -, \alpha_i \rangle$. But $\{\alpha_i\}_i$ is a basis, so ϕ therefore contradicts the nondegeneracy of $(-,-)$ on H^*. We conclude that A is in fact invertible, and because its entries are integers, we moreover conclude that A^{-1} have rational entries. Here we mean rational in the sense that F contains a subfield Q isomorphic to the usual rational numbers (the existence of this subfield follows from F having characteristic zero). Since this means that the coefficient vector $c = A^{-1}b$ of β has rational entries and since β was arbitrary, we have

$$\Phi \subset \operatorname{span}_Q \{\alpha_i\}_i. \tag{6.12}$$

After recalling that $\{\alpha_i\}_i \subset \Phi$, we see that we may write $\operatorname{span}_Q \Phi = \operatorname{span}_Q \{\alpha_i\}_i$. Since the $\{\alpha\}_i$ are linearly independent over F they certainly have to be over Q; hence $\operatorname{span}_Q \Phi$ has dimension ℓ when viewed as a vector space over Q.

This result certainly seems to suggest that sums of roots could be tractable to study–after all, we only need a few select roots and the rational scalars of F to be able to describe every other root. We understand well scalar multiples of a root, so let instead $\alpha, \beta \in \Phi$ be nonproportional: $\beta \neq \pm\alpha$. There can only be a finite number of roots of the form $\beta + i\alpha$, $(i \in \mathbb{Z})$, if any, and none of the spaces $L_{\beta+i\alpha}$ can equal H by assumption. By Proposition 6.4(i),

$$\begin{aligned} h_\alpha . L_{\beta+i\alpha} &\subset L_{\beta+i\alpha}, \\ x_\alpha . L_{\beta+i\alpha} &\subset L_{\beta+(i+1)\alpha}, \\ y_\alpha . L_{\beta+\alpha} &\subset L_{\beta+(i-1)\alpha}, \quad (i \in \mathbb{Z}), \end{aligned}$$

so we see that if we define

$$K = \bigoplus_{i \in \mathbb{Z}} L_{\beta+i\alpha}, \tag{6.13}$$

then K is an S_α-submodule of L under the adjoint action, and that all but a finite number of summands are zero. The set of weights of h_α in K is a subset of (6.6) and hence integers. More specifically,

$$\begin{aligned} \operatorname{ad}_L h_\alpha(x) = [h_\alpha x] &= (\beta + i\alpha)(h_\alpha)x \\ &= (\beta(h_\alpha) + i\alpha(h_\alpha))x = (\langle \beta, \alpha \rangle + 2i)x, \quad (x \in L_{\beta+i\alpha}), \end{aligned}$$

so that these weights are exactly $\langle\beta,\alpha\rangle+2i$ for those $i\in\mathbb{Z}$ such that $\beta+i\alpha\in\Phi$. It is clear that this set of numbers contains either exactly one of 0 and 1 or neither. In other words at most one of the weight spaces M_0 and M_1 is nonzero. Of course, the weight space corresponding to the weight $\langle\beta,\alpha\rangle+2i$ is exactly $L_{\beta+i\alpha}$ and is one-dimensional according to 6.10(a). We therefore have

$$\dim M_0 + \dim M_1 = 0 \text{ or } 1$$

and this formula is known from Corollary 8.6.1(c) to count the number of summands in any decomposition of (6.13) into irreducible submodules. *Hence K is irreducible* (it is nonzero since $L_\beta \subset K$) and Theorem 8.6 tells us that the previously described weights must form a "string"

$$\langle\beta,\alpha\rangle-2r,\quad \langle\beta,\alpha\rangle-2r+2,\quad \ldots,\quad \langle\beta,\alpha\rangle+2q-2,\quad \langle\beta,\alpha\rangle+2q,$$

for some pair of non-negative integers r,q. (To see this, observe that $\langle\beta,\alpha\rangle$ is a weight and must be in the string, so r,q just specifies how many steps the string takes in either direction.) Furthermore, the theorem tells us that the leftmost weight is the negative of the rightmost, the second leftmost is the negative of the second rightmost, and so on. It follows that $r-q=\langle\beta,\alpha\rangle$ by equating one endpoint with its negative. The absolute value of this number describes the "imbalance" of the string, i.e. how many more roots $\beta+i\alpha$ have $i>0$ rather than $i<0$ (or vice versa), and the sign describes in what direction the imbalance lies (e.g. a negative value means there are $|r-q|$ more roots in the "positive" direction). Since $\langle\beta,\alpha\rangle$ is a weight, we also obtain that $-\langle\beta,\alpha\rangle$ *is a weight, $L_{\beta-\langle\beta,\alpha\rangle\alpha}$ is a root space, and $\beta-\langle\beta,\alpha\rangle\alpha$ is a root.* This is our first result on roots and root spaces that decisively states how to construct new roots from old other than taking their negatives. Note that it technically holds even when $\beta=\pm\alpha$, since we just get $\beta-\langle\beta,\alpha\rangle\alpha=-\alpha$. We summarize these results in the next proposition.

Proposition 6.11. *Let $\alpha,\beta\in\Phi$ be roots. Then*

*(a) if $\beta\neq\pm\alpha$, then the set of roots that can be written as $\beta+i\alpha$, $(i\in\mathbb{Z})$ form an **α-string through β***

$$\beta-r\alpha,\quad \beta-(r-1)\alpha,\quad \ldots,\quad \beta+(q-1)\alpha,\quad \beta+q\alpha$$

for some non-negative integers r,q that satisfy $r-q=\langle\beta,\alpha\rangle$;

(b) $\beta-\langle\beta,\alpha\rangle\alpha\in\Phi$;

(c) if $\alpha+\beta\in\Phi$ then $[L_\alpha L_\beta]=L_{\alpha+\beta}$;

(d) L is generated by its root spaces, i.e. closing the set $\bigcup_{\alpha\in\Phi}L_\alpha$ under linear combinations and brackets yields exactly L.

Proof. (a) and (b) have already been shown.

(c) Lemma 8.5(b) says that the action of x_α on any irreducible S_α-module maps the one-dimensional weight space of weight λ to zero if λ is the maximal weight and onto the one-dimensional weight space of weight $\lambda+2$ otherwise. In particular, L_α maps L_β (of weight $\langle\beta,\alpha\rangle$) onto $L_{\beta+\alpha}$ (of weight $\langle\beta,\alpha\rangle+2$), which is equivalent to the statement since the action is the adjoint action.

(d) We can generate H one subspace at a time by taking $[L_\alpha L_{-\alpha}] = Ft_\alpha$, and then the assertion just follows from the Cartan decomposition.

□

Let us unpack part (b) of the proposition. It says that for any pair of roots $\alpha,\beta\in\Phi$ we have a root

$$\beta - 2\frac{(\beta,\alpha)}{(\alpha,\alpha)}\alpha \in \Phi, \tag{6.14}$$

in which $(-,-) : H^* \to F$ is the nondegenerate symmetric bilinear form we "borrowed" from the restriction of the Killing form to H. The above expression lends itself naturally to defining a linear operator $\sigma_\alpha : H^* \to H^*$:

$$\sigma_\alpha(\phi) := \phi - 2\frac{(\phi,\alpha)}{(\alpha,\alpha)}\alpha, \quad (\phi\in H^*). \tag{6.15}$$

This operator maps Φ into itself by (6.14). In fact, it is onto (i.e. $\sigma_\alpha(\Phi)=\Phi$) which follows from $\beta = \sigma_\alpha(\sigma_\alpha(\beta))$ and $\sigma_\alpha(\beta)\in\Phi$.
It turns out that σ_α admits a nice geometrical interpretation. Let E denote a euclidean space, by which we mean a finite-dimensional real vector space equipped with an inner product. Since the field is $\mathbb{R}$, an inner product here just refers to a positive-definite symmetric bilinear form on E. Though a slight abuse of notation, we denote the inner product of two vectors $u,v\in E$ by (u,v). To be positive-definite is to satisfy $(v,v)\geq 0$ for all $v\in E$ with equality if and only if $v=0$. Note that inner products are always nondegenerate: $(v,E)=0$ implies $(v,v)=0$ implies $v=0$. Fix nonzero $v\in E$. We define the *hyperplane determined by* v as the subspace

$$\Pi_v = \{u\in E \mid (u,v)=0\} = (\mathbb{R}v)^\perp.$$

Positive-definiteness implies $v\notin\Pi_v$, and it follows by nondegeneracy that the hyperplane has dimension $\dim E - 1$. In other words $E = \Pi_v\oplus\mathbb{R}v$ and we may therefore define a unique linear operator on E—the *reflection in* Π_v—by considering the operator that acts as the identity on Π_v and sends v to its negative. Write refl_v for this operator. An explicit formula is given by

$$\mathrm{refl}_v(u) := u - 2\frac{(u,v)}{(v,v)}v, \quad (u\in E) \tag{6.16}$$

which is due to the fact that this formula has exactly the two properties required in the definition, as is easily verified. Now compare (6.15) and (6.16) and note the similarities of the formulas. It would seem that roots behave as if they where a finite set of nonzero real vectors equipped with certain reflective symmetries—more specifically, one may reflect Φ in the "hyperplane" determined by any fixed root and again obtain Φ. If we allow ourselves some wishful thinking, we might imagine that there exists some way to canonically identify Φ with a geometrical object of the following type.

Definition 6.2. Let Φ be a subset of some euclidean space E and let $\ell = \dim E$. We say that Φ is a **root system of rank ℓ in E**, given that

(R1) $0 \notin \Phi$; Φ is finite; Φ spans E;

(R2) if $\alpha \in \Phi$ then $-\alpha \in \Phi$, but no other scalar multiple is;

(R3) $\langle \beta, \alpha \rangle := 2(\beta, \alpha)/(\alpha, \alpha) \in \mathbb{Z}$ for all $\alpha, \beta \in \Phi$.

(R4) for all $\alpha \in \Phi$, the reflection

$$\sigma_\alpha(\lambda) := \lambda - \langle \lambda, \alpha \rangle \alpha, \quad (\lambda \in E)$$

maps Φ onto itself.

Our earlier results on the Q-span of roots provide a good starting point for turning this intuition into reality. Recall that any vector space over F can be viewed as a vector space over the subfield $Q \subset F$ by only allowing scalar multiplication with rational scalars. (The canonical example is F_Q, which also illustrates that vector spaces constructed in this way need not have the same dimension as their parent space: e.g. $\mathbb{R}$ over $\mathbb{R}$ is one-dimensional but $\mathbb{R}$ over $\mathbb{Q}$ is infinite-dimensional, since the $\mathbb{Q}$-span of a finite set of real numbers can never cover all of $\mathbb{R}$.) With this in mind, let $E_Q = \operatorname{span}_Q \Phi$. It is by construction a subspace of H^* over Q, and our previously derived result (6.12) tells us that $\dim_Q E_Q = \ell = \dim_F H^*$, so the rank of Φ is exactly the (Q-)dimension of E_Q. We would like to bring along our form $(-,-)$ to E_Q and for this we have to show that it is rational-valued on E_Q. It clearly suffices to show that it has this property on Φ, which we do next.
Let $\phi, \phi' \in H^*$ be arbitrary. To these correspond unique elements $t_\phi, t_{\phi'} \in H$. Recall from our construction of the Cartan decomposition that $\operatorname{ad}_L t_\phi$, $\operatorname{ad}_L t_{\phi'}$ are simultaneously diagonalizable—more specifically, we should select a basis of each summand in (6.1) and then combine these bases, after which we obtain a basis relative to which $\operatorname{ad}_L t_\phi$, $\operatorname{ad}_L t_{\phi'}$ are both diagonal. In view of Proposition 6.10(a), we need only take a basis of H along with an arbitrary nonzero $x_\alpha \in L_\alpha$ for each $\alpha \in \Phi$. Now, $\operatorname{ad}_L t_\phi$ kills H (Lemma 6.1) and $\operatorname{ad}_L t_\phi(x_\alpha) = \alpha(t_\phi)x_\alpha$ for all $\alpha \in \Phi$, so

$$\operatorname{tr}(\operatorname{ad}_L t_\phi) = \underbrace{0 + \cdots + 0}_{\ell} + \sum_{\alpha \in \Phi} \alpha(t_\phi)$$

and similarly for $\operatorname{ad}_L t_{\phi'}$. Since both are diagonal relative to this basis,

$$\operatorname{tr}(\operatorname{ad}_L t_\phi \operatorname{ad}_L t_{\phi'}) = \sum_{\alpha \in \Phi} \alpha(t_\phi)\alpha(t_{\phi'}).$$

The left-hand side is by definition the Killing form applied to $t_\phi, t_{\phi'}$. We may now use the definition of $(-,-)$ to obtain the identity

$$\begin{aligned}(\phi,\phi') = \kappa(t_\phi, t_{\phi'}) &= \sum_{\alpha\in\Phi} \alpha(t_\phi)\alpha(t_{\phi'}) \\ &= \sum_{\alpha\in\Phi} \kappa(t_\alpha, t_\phi)\kappa(t_\alpha, t_{\phi'}) = \sum_{\alpha\in\Phi} (\alpha,\phi)(\alpha,\phi').\end{aligned}$$

In particular, taking $\phi = \phi'$ in the above yields

$$(\phi,\phi) = \sum_{\alpha\in\Phi} (\alpha,\phi)^2. \tag{6.17}$$

Fix $\beta \in \Phi$, let $\phi = \beta$, and divide both sides of (6.17) by $(\beta,\beta)^2$ to see that

$$\frac{1}{(\beta,\beta)} = \sum_{\alpha\in\Phi} \frac{(\alpha,\beta)^2}{(\beta,\beta)^2} = \frac{1}{4}\sum_{\alpha\in\Phi} \langle\alpha,\beta\rangle^2 \in Q.$$

Hence, $(\beta,\beta) \in Q$. We conclude that for all $\alpha, \beta \in \Phi$,

$$(\alpha,\beta) = \frac{1}{2}(\beta,\beta)\langle\alpha,\beta\rangle \in Q,$$

which is exactly what we wanted to show.
We now return to E_Q with the knowledge that that the restriction of $(-,-)$ to this Q-space is a well-defined symmetric bilinear form on E_Q. Write $(-,-)_Q$ for this restriction. It is a pleasant surprise that $(-,-)_Q$ is in fact *positive-definite*; just apply identity (6.17) to arbitrary $\lambda \in E_Q$ and note that the right-hand side is a sum of squares of rational numbers and has to be strictly positive unless $\lambda = 0$. With this we have done away with all dependencies on the original field and also have what amounts to a "Q-inner product", so all that remains is to exhibit a formal way to embed Φ into some euclidean space in such a way that $(-,-)_Q$ also embeds into the inner product on the space. This we do in the next theorem. We do not assume the readed is familiar with tensor products, so the construction of E is slightly complicated.

Theorem 6.12. *Let L be semisimple with given maximal toral subalgebra H. Let Φ be the corresponding roots. Construct $E = \mathbb{R} \otimes_Q E_Q$; then E is a real vector space, $(-,-)_Q$ induces an inner product $(-,-)$ on E, and Φ may be identified with a root system of rank $\dim_F H$ in E.*

For a definition of the tensor product, see Section 10.4 of [3].

Proof. Observe first that the axioms (R1)-(R4) all hold if one replaces E with E_Q and $(-,-)$ with $(-,-)_Q$ due to our having proved these at one point or another for Φ in the "larger" vector space H^*. Let $\ell = \dim_Q E_Q = \dim_F H$.

Take G to be the abelian group of all *formal finite sums of elements of* $\mathbb{R} \times E_Q$, i.e. all expressions of the form

$$n_1(r_1,\lambda_1) + \cdots + n_t(r_t,\lambda_t), \quad (t \in \mathbb{N}, n_i \in \mathbb{Z}, r_i \in \mathbb{R}, \lambda_i \in E_Q). \tag{6.18}$$

Each pair should be distinct but the order in which they appear do not matter; we equate two expressions if they are of the same length, contain the same pairs, and the corresponding integer coefficients are equal. We also include the case $t = 0$ as the empty sum. The group addition is given by "summing" the formal sums and combining multiple appearances of the same pair by summing their coefficients in $\mathbb{Z}$—we do not actually allow ourselves to sum up pairs "componentwise". It is not difficult to see that this indeed is an abelian group having the empty sum as the identity, and sums with coefficients of opposite sign as inverses. Next, define a subgroup H of G by including in it all formal sums

$$\begin{gathered}(r+s,\lambda)-(r,\lambda)-(s,\lambda)\\ (r,\lambda+\mu)-(r,\lambda)-(r,\mu)\\ (rq,\lambda)-(r,q\lambda),\end{gathered} \tag{6.19}$$

where $r, s \in \mathbb{R}$; $q \in Q$; $\lambda, \mu \in E_Q$, and then closing it under addition and inverses. (This guarantees that H is a subgroup.) Note that we have identified Q with $\mathbb{Q} \subset \mathbb{R}$, so rq makes sense. Any subgroup of an abelian subgroup is normal, so G/H is well-defined. The elements of G/H (which are cosets) are called *tensors*. A tensor that can be represented by single pair (r, λ) (i.e. this pair lies in the tensor/coset) is called *simple*, and is denoted $r \otimes \lambda$. It is important to point out that "most" tensors are not simple, but that we nevertheless can write them (non-uniquely in general) as a sum of simple tensors—for example, the tensor represented by (6.18) can be decomposed as

$$n_1(r_1 \otimes \lambda_1) + \cdots + n_t(r_t \otimes \lambda_t). \tag{6.20}$$

By definition of cosets, $r \otimes \lambda = s \otimes \mu$ if and only if $(r, \lambda) - (s, \mu) \in H$, so the fact that H contains all elements of the form (6.19) translates to

$$\begin{gathered}(r+s)\otimes\lambda = r\otimes\lambda + s\otimes\lambda\\ r\otimes(\lambda+\mu) = r\otimes\lambda + r\otimes\mu\\ (rq)\otimes\lambda = r\otimes(q\lambda)\end{gathered}$$

for all r, s, q, λ, μ as above. This tells us why (6.20) need not be unique.

Next, we turn G/H into a vector space over $\mathbb{R}$ by defining scalar multiplication in terms of simple tensors:

$$r.\left(\sum_{i=1}^{n}(s_i \otimes \lambda_i)\right) = \sum_{i=1}^{n}((rs_i) \otimes \lambda_i), \quad (n \in \mathbb{Z}^+; r, s_i \in \mathbb{R}; \lambda_i \in E_Q).$$

One can verify that this is well-defined (see Section 10.4 of [3]) and that the resulting structure satisfies the axioms required of a vector space. By definition, $E = \mathbb{R} \otimes_Q E_Q = G/H$ and is called their *tensor product* over Q. The corresponding vector space is called the *scalar extension of E_Q to $\mathbb{R}$*.

Let $\phi : E_Q \to \mathbb{R}^\ell$ be the Q-isomorphism given by $\phi(\sum q_i\alpha_i) = \sum q_i e_i$, where $\{\alpha_i\}_i \subset \Phi$ is an F-basis of H^* and $\{e_i\}_i$ is the standard basis of $\mathbb{R}^\ell$. Let

$i : E_Q \to E$ be the map $i(\lambda) = 1 \otimes \lambda$. In the section referenced above, i is shown to be Q-linear. Moreover, Theorem 8 of this section says that there exists a unique $\mathbb{R}$-linear map $\varphi : E \to \mathbb{R}^\ell$ such that the following diagram commutes.

$$\begin{array}{ccc} E_Q & \xrightarrow{\;i\;} & E \\ & \underset{\phi}{\searrow} & \downarrow{\scriptstyle \varphi} \\ & & \mathbb{R}^\ell \end{array}$$

Since ϕ is bijective, i must be injective and φ must be surjective. As remarked in the proof of the same theorem, E is the $\mathbb{R}$-span of simple tensors of the form $1 \otimes \lambda$, $(\lambda \in E_Q)$, i.e. E is the $\mathbb{R}$-span of $i(E_Q)$. Now, $\{\alpha_1, \ldots, \alpha_\ell\}$ is a basis of E_Q, meaning E is the $\mathbb{R}$-span of $\{i(\alpha_i)\}_i$. Also, E has to have Q-dimension at least ℓ for i to be injective. We conclude that $\{i(\alpha_1), \ldots, i(\alpha_\ell)\}$ *is an* $\mathbb{R}$*-basis of* E. In particular, $\dim E = \ell$.

Finally, define a bilinear form $(-,-)$ on E by letting

$$(i(\alpha_n), i(\alpha_m)) = (\alpha_n, \alpha_m)_Q, \quad (1 \le n, m \le \ell)$$

and then extending it $\mathbb{R}$-linearly to all of E. Symmetry immediately follows from that of $(-,-)_Q$. One can see that $(-,-)$ is positive-definite by considering its corresponding quadratic form; the matrix of this form is just the matrix of the quadratic form of $(-,-)_Q$, which we already know is positive-definite. Hence, $(-,-)$ is an inner product on E. Its definition along with the Q-linearity of i also guarantees that $(i(\alpha), i(\beta)) = (\alpha, \beta)_Q$ for all roots $\alpha, \beta \in \Phi$. It is then clear that $i(\Phi)$ satisfies (R1)-(R4) due to the remark in the beginning of this proof. Therefore Φ is Q-isomorphic to a root system—namely $i(\Phi)$—of rank ℓ in E.

□

7 Summary and further discussion

Lie algebras arise naturally as the tangent space at the identity of so-called Lie groups. Although such algebras are real vector spaces, we prefer to work with Lie algebras over algebraically closed fields as they are more receptive to certain tools from linear algebra. To study these effectively, one defines both "abstract" and "concrete" (linear) Lie algebras, and shows how to canonically inject each abstract Lie algebra into a concrete one (by means of the adjoint representation). This allows us to, among other things

- define and investigate properties of abstract Lie algebras by "transferring" known results on concrete algebras across this injection, and
- to prove results in their greatest generality, by proving them for abstract algebras when possible, or else inject a concrete algebra into a concrete one and exploit the multitude of relationships this brings about.

Once we have used this to derive a satisfyingly large body of results for Lie algebras in general, we then specialize to those can be built up from atomic ones (the semisimple algebras). After learning that all such algebras must contain subalgebras of a certain type, we use this to our advantage to systematically decompose them into subspaces (the Cartan decomposition). Finally, we find that to any given such decomposition is associated a rigid, geometric object that can be described in but a few axioms, and that this object contains essential information about the decomposition, and in turn the algebra itself.

It is at this point that we end our inquiry into Lie algebras. It would however be unreasonable to expect the reader to see the value of all our specialized theory without suggesting what it all leads up to. The following list is meant as an informal discussion on some further results in the subject, and the author does not make any claims on having a perfect understanding of these results and/or how to show them.

- There is a natural definition of isomorphisms between and in particular automorphisms of root systems.
- By choosing $\{\alpha_i\}_i$ more carefully, one can sharpen (6.12) to the case where each root can be written with integer coefficients, all nonnegative or non-positive. This is called a *base* of Φ.
- Any root system is obtained as a union of "irreducible" systems.
- Simple algebras have irreducible root systems; semisimple algebras have root systems built up from the irreducible root systems of its simple ideals.
- *Irreducible root systems admit a complete classification, and each can be explicitly constructed.*
- The notion of maximal toral subalgebra can be generalized, and this allows one to prove that *the root system of a semisimple algebra is independent of the choice of maximal toral subalgebra.*
- Each root system is the root system of some semisimple Lie algebra.
- Every isomorphism of two root systems, sending a base to a base (along with certain other specifications), extends uniquely to an isomorphism of the corresponding semisimple Lie algebras.

With this, the author would like to extend his gratitude to the reader for lending their attention to his thesis.

References

[1] S. Axler. *Linear Algebra Done Right, 2nd ed.* Springer-Verlag, 1997.

[2] K. Conrad. Simultaneous commutativity of operators. `http://www.math.uconn.edu/~kconrad/blurbs/linmultialg/simulcomm.pdf`. [Online; accessed 9-May-2018].

[3] D. S. Dummit and R. M. Foote. *Abstract Algebra, 3rd ed.* John Wiley and Sons, Inc., 2004.

[4] W. Fulton and J. Harris. *Representation Theory: A First Course*. Springer Science+Business Media, Inc., 2004.

[5] B. C. Hall. *Lie Groups, Lie Algebras, and Representations: An elementary Introduction*. Springer Science+Business Media New York, 2003.

[6] A. Holst and V. Ufnarovski. *Matrix Theory*. Studentlitteratur AB, 2014.

[7] J. E. Humphreys. *Introduction to Lie Algebras and Representation Theory*. Springer-Verlag, 1972.

8 Appendix

Here we describe some results that are necessary only for certain parts of the theory and which can proven somewhat independently. The first section can be safely skipped, but the second introduces the concept of weights, which play an important role in understanding the Cartan decomposition.

8.1 Theorems on nilpotent and solvable algebras

In the first of these theorems, the underlying field F may be arbitrary. In the second we require char $F = 0$.

Theorem 8.1. *Let $L \subset \mathfrak{gl}(V)$ be a linear Lie algebra, where V is finite dimensional and nonzero. Suppose every operator in L is nilpotent. Then L has a common eigenvector with eigenvalue 0, i.e. $L.v = 0$ for some nonzero $v \in V$.*

Proof. If $\dim L = 0$ then any nonzero $v \in V$ will do. If $\dim L = 1$ then L is the span of some nonzero (nilpotent) operator $x \in L$. This operator always has at least one eigenvector with eigenvalue 0, namely the last nonzero vector of the sequence $x^n.w$ for any fixed nonzero $w \in V$, and this serves as our v. Now let $\dim L \geq 2$. Take as induction hypothesis that the theorem is true for all $M \subset \mathfrak{gl}(U)$ that satisfies the conditions of the theorem along with the condition $\dim M < \dim L$. We prove the remainder of the theorem in steps.

Let K be a maximal proper subalgebra of L, and let $W = \{v \in V \mid K.v = 0\}$. The latter is a subspace, and moreover nonempty since $K \subset \mathfrak{gl}(V)$ satisfies the induction hypothesis.

(1) $0 < \dim K < \dim L$.

The span of any nonzero $x \in L$ is a one-dimensional subalgebra of L, with trivial bracket. By $\dim L \geq 2$ such subalgebras exists and are proper. If no larger proper subalgebras exist we may thus take K to be one of these.

(2) K is properly included in $N_L(K)$.

Consider the representation $\phi : K \to \mathfrak{gl}(L/K)$ as constructed in the final remark of Section 2.2 (in the notation used there, $\phi = (\mathrm{ad}_L|_K)'$). Take $M = \mathrm{Im}\, \phi$ and $U = L/K$. Observe that $\dim M \leq \dim K < \dim L$, and $U \neq 0$ by (1). Any $x \in K$ is nilpotent by hypothesis, so $\mathrm{ad}_L\, x$ is (Lemma 3.1(i)), and then so is $\phi(x)$ by using the identity $\phi(x)(y + K) = \mathrm{ad}_L\, x(y) + K$. Hence $M \subset \mathfrak{gl}(U)$ satisfies the induction hypothesis, so there exists $y + K \in L/K$, $y + K \neq K$ (i.e. nonzero) such that $L.(y + K) = 0$. Equivalently, $y \notin K$ and $[Ly] \subset K$. Then $y \in N_L(K) \setminus K$, and y is a fortiori nonzero since $y \notin K$.

(3) K is an ideal of L.

Since K is maximal, (2) forces $N_L(K) = L$.

(4) $L = K \oplus Fz$, $(z \in L \setminus K)$.

First, L/K is a Lie algebra by (3). Suppose $\dim L/K > 1$; reasoning as in (1) we take a (proper) one-dimensional subalgebra of L/K, and the preimage of this subalgebra along the quotient homomorphism $\pi : L \to L/K$ yields a proper subalgebra of L that properly contains K. This contradicts the maximality of K, so $\dim L/K = 1$. Equivalently, $\dim L = \dim K + 1$, and since $K + Fz$ has higher dimension than K we must have $K \oplus Fz = L$.

(5) $z.v = 0$ for some nonzero $v \in W$.

Choose z that satisfies (4). Let $w \in L.W$, that is, $w = x.v$ for some $x \in L$, $v \in W$. Then (3) and axiom (M3) imply

$$K.w = K.(x.v) = x.(K.v) - [xK].v \subset x.0 - K.v = 0.$$

Hence $w \in W$, showing that $L.W \subset W$. Take $M = Fz$ and $U = W$. In particular, $M.U \subset U$. Then $M \subset \mathfrak{gl}(U)$ satisfies the induction hypothesis, so there exists nonzero $v \in W$ such that $M.v = 0$, or equivalently $z.v = 0$.

(6) $L.v = 0$.

By (4) and (5), $L.v = (K \oplus Fz).v = K.v + F(z.v) = 0$. The theorem now follows for all $\dim L = 0, 1, 2, \ldots$ by induction.

□

Let $L \subset \mathfrak{gl}(V)$ be a linear Lie algebra and let $(x_1, \ldots, x_n)$ be a basis of L. Suppose that $v \in V$ is an eigenvector of each basis element, meaning $x_i.v = \lambda_i v$ for some $\lambda_i \in F$, $1 \leq i \leq n$. Define a linear functional $\lambda : L \to F$ by $x_i \mapsto \lambda_i$ and then extending it linearly to all of L. This implies that $x.v = \lambda(x)v$ for all $x \in L$, so v is an eigenvector of every $x \in L$ with corresponding eigenvalue $\lambda(x)$. We call such a pair (v, λ) a *common eigenvector and eigenvalue* of L. The statement that such a pair exists is equivalent to the statement that the space

$$V_{L,\lambda} := \{w \in V \mid x.w = \lambda(x)w \text{ for all } x \in L\}$$

is nonempty for some choice of $\lambda \in L^*$.

Theorem 8.2. *Let $L \subset \mathfrak{gl}(V)$ be a linear Lie algebra, where V is finite dimensional and nonzero. Suppose L is solvable. Then L has a common eigenvector in V, i.e. there exists a nonzero vector $v \in V$ along with a linear functional $\lambda \in L^*$ such that $x.v = \lambda(x)v$ for all $x \in L$.*

Proof. If $\dim L = 0$ (i.e. $L^* = 0$) then any nonzero $v \in V$ will do, so suppose $\dim L \geq 1$. Take as induction hypothesis that the theorem holds for any solvable algebra $K \subset \mathfrak{gl}(V)$ of dimension less than L.

(1) L has an ideal K such that $L = K \oplus Fz$, $(z \in L \setminus K)$.

It is necessary that $[LL] \neq L$ for L to be solvable in view of our assumption that L has positive dimension. Hence $L/[LL]$ is nonzero and has a subspace

$K/[LL]$ of dimension one less than $L/[LL]$. As we showed in Section 2, the latter is abelian, which means that $K/[LL]$ is an ideal. Then K is an ideal of L by Theorem [] and $\dim L - \dim K = \dim L/[LL] - \dim K/[LL] = 1$ shows that is has the correct dimension.

(2) $V_{K,\mu}$ is nonempty for some $\mu \in K^*$.

Here we use the induction hypothesis: K is solvable by Proposition 4.4(a) and clearly satisfies the other conditions of the theorem, so we obtain a nonzero vector $w \in V$ along with a linear functional $\mu \in K^*$ such that $y.w = \mu(y)w$ for all $y \in K$. This w lies in $V_{K,\mu}$ by definition.

(3) For any given $x \in L$ exists a nonzero subspace $W \subset V$ such that $x.W \subset W$ and $K.W \subset W$.

Fix nonzero $w \in V_{K,\mu}$ (which is possible by the above) and consider the sequence of subspaces

$$W_0 := 0, \quad W_i := \operatorname{span}_F\{w, x.w, \dots, x^{i-1}.w\}, \quad i \geq 1.$$

Each W_i, $i \geq 1$ is nonzero and $x.W_i \subset W_{i+1}$ for all $i \geq 0$. Observe also that $K.W_i \subset W_i$ by K being an ideal—we can see this by looking at how $y \in K$ acts on the given spanning set of W_i, $(i \geq 1)$:

$$y.(x^j.w) = [yx^j].w + x^j.(y.w) = \mu([yx^j])w + \mu(y)x^j.w, \quad 0 \leq j \leq i-1.$$

V is finite-dimensional so there must exist a positive integer n such that

$$\dim W_1 = 1, \quad \dots, \quad \dim W_n = n, \quad \dim W_{n+1} = n, \quad \dots$$

for which we then let $W := W_n$. This choice guarantees $x.W_n \subset W_n$, and it follows from our earlier considerations that W has the desired properties.

(4) W has a basis relative to which each $y|_W$, $(y \in K)$ is represented by an upper triangular matrix with $\mu(y)$ on the diagonal.

Let $y \in K$; $y|_W$ is well-defined because $y.W \subset W$. The basis in question is of course $(w, x.w, \dots, x^{n-1}.w)$. Write $w_i = x^{i-1}.w$. With this basis, the rest of the statement follows if we can show that

$$\text{for each } i \geq 1 \text{ exists } w' \in W_{i-1} \text{ such that } y.w_i = \mu(y)w_i + w'. \tag{8.1}$$

We prove this by induction. Recall that $W_0 = 0$, so for $i = 1$ we have to choose $w' = 0$, but this works out because $y.w_1 = \mu(y)w_1$. Now suppose (8.1) is true for $i = k$. Using this along with $[xy].w_k \in K.W_k \subset W_k$ where the inclusion was obtained in (3), we see that

$$\begin{aligned} y.w_{k+1} &= y.(x.w_k) \\ &= x.(y.w_k) - [xy].w_k \\ &= x.(\lambda(y)w_k + w') - [xy].w_k \quad (w' \in W_{k-1}) \\ &= \lambda(y)x.w_k + x.w' - [xy].w_k \\ &= \lambda(y)w_{k+1} + w'' \end{aligned}$$

where $w'' = x.w' - [xy].w_k \in W_k$ since both summands lies in this subspace. Therefore (8.1) is true for $i = k + 1$, and for every $i \geq 1$ by induction.

(5) $\mu([L, K]) = 0$

Let $x \in L$, $y \in K$ be arbitrary and construct W as in (3). Since both x and y stabilize W, their commutator does, too, and therefore $[xy]|_W$ is well-defined. We may use the identities (3.2), (3.3) to verify that the trace of $[xy]|_W$ is zero. But $[xy] \in K$, so (4) says that this trace should equal $n\mu([xy])$ where $n = \dim W$. Invoking our assumption that char $F = 0$ (in particular, $n \neq 0$) allows us to conclude that $\mu([xy]) = 0$.

(6) $L.V_{K,\mu} \subset V_{K,\mu}$.

Let $x \in L$, $v \in V_{K,\mu}$ be arbitrary. For $x.v$ to lie in $V_{K,\mu}$ it should satisfy the defining condition of this subspace, which is

$$y.(x.v) = \mu(y)x.v \text{ for all } y \in K.$$

We can rewrite the left-hand side as $x.(y.v) - [xy].v = \mu(y)x.v - \mu([xy])v$ and then (5) immediately gives that $x.v$ indeed satisfies the condition.

(7) $V_{L,\lambda}$ is nonempty for some $\lambda \in L^*$.

Write $V_0 = V_{K,\mu}$ and use (1) to decompose $L = K \oplus Fz$ for some $z \in L \setminus K$. The restriction $z|_{V_0}$ is well-defined in view of (6). We can always find *some* eigenvector for any given linear operator of a vector space over F since the field is algebraically closed; hence $z|_{V_0}$ has an eigenvector $v \in V_0$ with some eigenvalue $\mu_0 \in F$. Let $x \in L$ and write it uniquely of the form $x = y + az$, $(y \in K, a \in F)$. Then

$$x.v = y.v + az.v = (\mu(y) + a\mu_0)v$$

which means that after defining a linear functional $\lambda \in L^*$ by

$$\lambda(x) = \mu(y) + a\mu_0$$

and similarly for all other $x \in L$, we obtain that $v \in V_{L,\lambda}$.

□

8.2 Representations of $\mathfrak{sl}(2, F)$

Let V be a finite-dimensional vector space over F. We denote by $\mathfrak{sl}(V)$ the set of all linear operators of V with vanishing trace. By identity (3.2) of Section 3.2, $\mathfrak{sl}(V)$ is a subspace of $\mathfrak{gl}(V)$. Using identity (3.3) of the same section we also obtain $\mathrm{tr}([xy]) = 0$, $x, y \in \mathfrak{gl}(V)$. In other words, $[\mathfrak{gl}(V), \mathfrak{gl}(V)] \subset \mathfrak{sl}(V)$, which in particular shows that $\mathfrak{sl}(V)$ is a subalgebra. Letting $n = \dim V$, we may identify $\mathfrak{sl}(V)$ with the set of $n \times n$ zero-trace matrices over F, which we denote

$\mathfrak{sl}(n, F)$. In this section we shall concern ourselves exclusively with $S = \mathfrak{sl}(2, F)$, to which we assign the standard basis

$$h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Due to linearity, the bracket of S is completely determined by its behaviour on these basis vectors, which we compute to be $[hx] = 2x$, $[hy] = -2y$, $[xy] = h$.

Lemma 8.3. *If* char $F \neq 2$, *then S is simple.*

Proof. We need to show that any nonzero ideal I equals L. Let such I be given, and pick nonzero $z = ah + bx + cy \in I$. If $b = c = 0$, then $h \in I$, in which case $[hx] = 2x \in I$, $[hy] = -2y \in I$. The hypothesis $2 \neq 0$ then implies $x, y \in I$, and we obtain $I = L$. Hence, suppose at least one of b, c is nonzero. We have

$$[x[xz]] = -2cx \in I, \quad [y[yz]] = -2by \in I,$$

so $b \neq 0$ ($c \neq 0$) implies $y \in I$ ($x \in I$). In either case, $[xy] = h \in I$, and we again conclude that $I = L$. □

In the sequel, assume char $F = 0$ and algebraic closedness. As suggested by the title of the section we now turn our attention to representations. Let V be a finite-dimensional S-module with corresponding representation $\phi : S \to \mathfrak{gl}(V)$. Given $\lambda \in F$, the subspace $V_\lambda = \{v \in V \mid h.v = \lambda v\}$ is nonzero if and only if λ is an eigenvalue of $\phi(h)$, since $\phi(h)(v) = h.v$. When this is the case we call λ a *weight* of h in V, and V_λ a *weight space*. Now, S is simple (in particular, semisimple), so Corollary 5.6.2 says that because h equals its semisimple part in S, so must $\phi(h)$ in $\mathfrak{gl}(V)$. Hence V has a basis of eigenvectors of $\phi(h)$, meaning *V is the direct sum of its weight spaces.* Naturally, we investigate how S acts on its weight spaces or more generally the spaces V_λ.

Lemma 8.4. *With x, y, h; V_λ, $\lambda \in F$ as above,*

(a) $h.V_\lambda \subset V_\lambda$;

(b) $x.V_\lambda \subset V_{\lambda+2}$;

(c) $y.V_\lambda \subset V_{\lambda-2}$.

Proof. (a) is immediate from the definition of V_λ. For (b), let $v \in V_\lambda$. Then

$$h.(x.v) = [hx].v + x.(h.v) = 2x.v + \lambda x.v = (\lambda + 2)x.v,$$

so we see that $x.v \in V_{\lambda+2}$. (c) follows in a similar fashion. □

The number of weight spaces is finite (V being finite-dimensional) so the lemma implies that some weight space V_μ must be killed by x—simply start with any weight space V_λ and take the last nonzero space of the sequence $x^i.V_\lambda \subset V_{\lambda+2i}$, $i \geq 0$. We call any nonzero vector of any such V_μ a *maximal vector* of weight

μ. The existence of maximal vectors provides us with a systematic way of constructing a nonzero submodule of V, in the following way. Fix a maximal vector $v_0 \in V_\mu$, and define $v_{-1} = 0$, $v_i = \frac{1}{i!} y^i.v_0$, $i \geq 1$. The action of S on these vectors is given by

Lemma 8.5. *With v_i, $i = -1, 0, 1, 2, \dots$ as above,*

(a) $h.v_i = (\mu - 2i)v_i$;

(b) $x.v_i = (\mu - i + 1)v_{i-1}$;

(c) $y.v_i = (i+1)v_{i+1}, \quad (i \geq 0)$.

Proof. By Lemma 8.4(c), $v_i \in V_{\mu-2i}$ for all $i \geq 0$, which proves (a). (c) follows from the definition of v_i. (b) is true for $i = 0$, and we use induction to prove it for higher i, along with a few applications of (a) and (c).

$$\begin{aligned}
(k+1)x.v_{k+1} &= x.\frac{1}{k!}y^{k+1}.v_0 = x.y.v_k \\
&= [xy].v_k + y.x.v_k = h.v_k + y.(x.v_k) \\
&= (\mu - 2k)v_k + (\mu - k + 1)y.v_{k-1} \\
&= (\mu - 2k)v_k + k(\mu - k + 1)v_k \\
&= (\mu + k\mu - k - k^2)v_k \\
&= (1+k)(\mu - k)v_k, \quad (k \geq 0)
\end{aligned}$$

The induction assumption that (b) holds for k was deployed in the fifth equality. Dividing by $k+1$ in the above yields $x.v_{k+1} = (\mu - k)v_k$, which proves that (b) holds for $i = k+1$, and we are done. □

Analogously to how x kills V_μ, y must eventually kill the sequence $V_{\mu-2i}$, $i \geq 0$. Take m to be the smallest i for which $V_{\mu-2i}$ is killed by y, but still is a weight space. Then $(v_0, \dots, v_m)$ are nonzero, and in addition linearly independent, due to their lying in different weight spaces. Lemma 8.5 says that the action of S either shift around and/or scale these vectors, or kills them, and therefore $W = \operatorname{span}\{v_0, \dots, v_m\}$ *is an S-submodule of V*. Of course, V may contain many possible choices of maximal vector, and thus many different submodules, but if V is irreducible, then W being nonzero forces $V = W$ (in particular, $\dim V = \dim W = m + 1$). Gathering the results so far:

Theorem 8.6. *Let V be an irreducible S-module. The weight spaces of V are*

$$V_{-m},\ V_{-m-2},\ \dots,\ V_{m-2},\ V_m \quad \textit{where} \quad m = \dim V - 1,$$

and each is one-dimensional. Moreover, V has a basis $(v_0, \dots, v_m)$, $v_i \in V_{m-2i}$ relative to which the standard basis h, x, y of S is represented by, respectively,

$$\begin{pmatrix} m & 0 & \dots & 0 \\ 0 & m\text{-}2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{-}m \end{pmatrix}, \quad \begin{pmatrix} 0 & m & 0 & \dots & 0 \\ 0 & 0 & m\text{-}1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & m & 0 \end{pmatrix}.$$

Proof. Construct W as above, so $V = W$ and $\dim V = m + 1$. We already know that $(v_0, \ldots, v_m)$ is a basis, that $v_i \in V_{\mu-2i}$, and that these weight spaces are disjoint, which forces them to be one-dimensional. That the standard basis of S is represented as claimed may be verified by translating the identities of Lemma 8.5—which determine the action of h, x, y on the chosen basis—into matrix form. Finally, applying Lemma 8.5(b) to v_{m+1}, which, of course, is identically zero, yields $0 = x.v_{m+1} = (\mu - m)v_m$. But v_m is nonzero, so we must have $\mu = m$. □

Any irreducible S-module V therefore has, up to multiplication by nonzero scalar multiples, a unique maximal vector with unique weight $m = \dim V - 1$. We call this non-negative integer the *maximal weight* of V. All other weights of V are generated by starting with m and subtracting 2 until we reach $-m$.
Having classified the irreducible S-modules, we now turn to arbitrary modules. Since S is simple (in particular, semisimple) we can apply Weyl's Theorem on complete reducibility to decompose any such module into irreducible submodules, and these are completely described by the theorem. The first two results of the next corollary are analogous to those that hold for irreducible submodules, while the third is new.

Corollary 8.6.1. *Let V be a finite-dimensional S-module, with corresponding representation $\phi : S \to \mathfrak{gl}(V)$, and let $V = W_1 \oplus \cdots \oplus W_t$ be a decomposition of V into irreducible S-submodules. Let $V_\lambda = \{v \in V \mid \phi(x)(v) = \lambda v\}$, i.e. $V_\lambda \neq 0$ if and only if λ is an eigenvalue of $\phi(h)$. Then*

(a) all eigenvalues of $\phi(h)$ are integers;

(b) $\dim V_m = \dim V_{-m}$ (in particular, m is an eigenvalue if and only if $-m$ is);

(c) $t = \dim V_0 + \dim V_1$, so the number of irreducible submodules present in any such decomposition of V is uniquely determined.

Proof.

(a) Fix nonzero $v \in V$ and let $w_i \in W_i$, $1 \leq i \leq t$ be the unique vectors such that $v = \sum w_i$. Because v is nonzero, some of the w_i has to be nonzero—say, $w_{i_1}, \ldots w_{i_k}$, $(1 \leq k \leq t)$. Now, v is an eigenvector of $\phi(h)$ with eigenvalue λ if and only if

$$\sum h.w_i = h.\left(\sum w_i\right) = h.v = \phi(h)(v) = \lambda v = \sum \lambda w_i,$$

which by $h.w_i \in W_i$ and the direct sum is true exactly when $h.w_i = \lambda w_i$ for all $1 \leq i \leq t$. In particular, $h.w_{i_j} = \lambda w_{i_j}$, meaning λ is a weight of h in W_{i_j} for all $1 \leq j \leq k$. But the theorem says that all of these weights are integers, so it follows that λ must be an integer.

(b) It is clear from the proof of (a) that without loss of generality, every eigenvector v of eigenvalue m has a unique representation $v = \sum_{i=1}^{k} w_i$ where each $w_i \in W_i$ is an eigenvector with eigenvalue m and $1, \ldots, k$ are exactly the indices of those irreducible submodules which has m as a weight. The theorem tells us that the weight space of m in W_i, $1 \leq i \leq k$ is one-dimensional, so there is up to scalar multiples only one possible choice for each w_i. Hence, $\dim V_m = k$. The theorem also tells us that $1, \ldots, k$ are exactly the indices of those irreducible submodules which has $-m$ as a weight, and it follows that $\dim V_{-m} = k = \dim V_m$.

(c) Let W be one of the irreducible submodules and write m for its maximal weight. By the theorem, the set of weights of h in W form an arithmetic progression with common difference 2 that moreover is symmetric around zero. This implies that exactly one of $m \equiv 0 \pmod 2$ or $m \equiv 1 \pmod 2$ is true. By slightly modifying the argument in (b), one obtains that $\dim V_0$ counts the number of irreducible submodules where the former is the case and $\dim V_1$ counts the latter. Together, these add up to the total number of irreducible submodules of the decomposition.

□

Interestingly, (c) says that one could prove that a given S-module is irreducible by showing that it has a unique (up to scalar multiples) eigenvector of eigenvalue either 0 or 1, but no eigenvectors of the other. This is because it then already is its own decomposition into irreducible submodules. Once this has been done, we then have a complete description of the module through Theorem 8.6.