Essays on representations of real groups

Introduction to Lie algebras

Bill Casselman
University of British Columbia
cass@math.ubc.ca

This essay is intended to be a self-contained if rather brief introduction to Lie algebras. The first several sections deal with Lie algebras as spaces of invariant vector fields on a Lie group. The very first sections are concerned with different ways to consider vector fields. This early treatment is intended, among other things, to motivate the eventual introduction of abstract Lie algebras, with which all the later sections are concerned.

The first part is not intended to be complete in all minor points, just to give a rough idea of how things go. The second part, on the contrary, is intended to be as complete as possible, but only treats results that lead without too much digression to the final goal of understanding semi-simple and reductive Lie algebras. In the third part I sketch the proof that semi-simple Lie algebras give rise to root systems, although I do not discuss the consequences of this. In particular, I do not say anything about the structure of root systems or results about semi-simple Lie algebras that depend on it, such as the structure of the centre of the enveloping algebra or the classification of irreducible representations. That is another story.

In much of this account, I have followed the standard references [Jacobson:1962], [Serre:1965], and [Serre:1966]. However, although nothing I do is without some basis already in the literature, I believe that some aspects of the proofs in the second part are new. One thing I’d like to point out is my use of primary decompositions of finite-dimensional modules over nilpotent Lie algebras, which I found in [Jacobson:1962]. I find it to be both interesting and illuminating, especially in the proof of Cartan’s criterion for solvability, where I find it to be the natural tool. I’m surprised one doesn’t come across it more often. Perhaps because of the role of the Jacquet and Whittaker functors in the theory of representations of real reductive Lie algebras, it deserves to be better known.

The amount of space allotted to various topics will probably look idiosyncratic to some. But for this I offer no explanation other than that I became interested in pursuing things often left out in other expositions of this material. The overall length of this essay is still rather less than that of other introductions, so I probably need to make no apology for what I hope to be amiable bavardage.

Contents

I. Lie algebras and vector fields
   1. Manifolds, bundles, sheaves
   2. Vector fields and differential operators
   3. The Lie algebra of a Lie group
   4. Normal coordinates
   5. The Lie algebra of SL(2,R)
   6. Vector fields associated to $G$-actions

II. Lie algebras on their own
   7. Abstract Lie algebras
   8. Representations
   9. Nilpotent Lie algebras
  10. Representations of a nilpotent Lie algebra
  11. Cartan subalgebras
  12. Conjugacy of Cartan subalgebras
Part I. Lie algebras and vector fields

1. Manifolds, bundles, sheaves

Let $M$ be a manifold. The assignment of $C^\infty(U) = C^\infty(U, \mathbb{R})$ to every open subset $U$ of $M$ is a sheaf. That is to say, if $\{U_\alpha\}$ is a covering of $U$, then there exists a sequence that is exact in a sense I’ll explain in a moment:

$$0 \longrightarrow C^\infty(U) \longrightarrow \prod_\alpha C^\infty(U_\alpha) \longrightarrow \bigoplus_{\beta, \gamma} C^\infty(U_\beta \cap U_\gamma).$$

The double sum is over all ordered pairs $\beta, \gamma$. The top arrow maps $C^\infty(U_\alpha)$ by restriction to factors $C^\infty(U_\alpha \cap U_\gamma)$, and by 0 to other factors; the second similarly to $\bigoplus_\beta C^\infty(U_\beta \cap U_\alpha)$. The sequence is exact in the sense that $C^\infty(U)$ is the subset of $\prod_\alpha C^\infty(U_\alpha)$ on which both maps agree. Informally, the sheaf property encapsulates the notion that being a smooth function is a local property—a function is smooth on an open set $U$ if and only if it is smooth in the neighbourhood of every one of its points.

To a sheaf is associated its stalks at every point. The stalk of the sheaf associated to $C^\infty(U)$ is the local ring $C_p$ of germs of smooth functions in the neighbourhood of a $p$. Such a germ is a smooth function defined in the neighbourhood of $p$, but modulo the relation that two such functions are equivalent if the agree in some (likely smaller) neighbourhood of $p$. In other words, $C_p$ is the direct limit of the spaces $C^\infty(U_\alpha)$ as $\alpha$ ranges over the open neighbourhoods of $p$. Every function smooth in the neighbourhood of $p$ determines an element of $C_p$ and, as I have said, two functions defined in two neighbourhoods determine the same element if they agree on some neighbourhood of $p$ contained in both of them. In other words, the image in $C_p$ of a function in some $C^\infty(U)$ is determined completely by its image in arbitrarily small neighbourhoods of $p$.

1.1. Lemma. The map from $C^\infty(M)$ to $C_p$ is surjective.

Proof. Suppose $U$ to be a neighbourhood of $p$. Given $f$ in $C^\infty(U)$, one may choose a smooth function $\varphi$ of compact support contained in $U$, identically 1 near $p$. Then $\varphi f$ lies in $C^\infty(M)$ and agrees with $f$ near $p$.  

The following is a modest improvement of a familiar assertion about the Taylor series of a smooth function:

1.2. Lemma. Suppose $U$ to be a convex open neighbourhood of the origin in $\mathbb{R}^n$. If $f$ is a smooth function on $U$ then for any $m$ it may be expressed as

$$f(x) = \sum_{|k|<m} f^{(k)}(0) \frac{x^k}{k!} + \sum_{|k|=m} x^k f_k(x)$$

where each $f_k$ is a smooth function on $U$.

Here $k = (k_i)$ is a multi-index:

$$|k| = \sum k_i, \quad x^k = \prod x_i^{k_i}, \quad k! = \prod k_i! \quad \text{and} \quad f^{(k)} = \frac{\partial^{|k|}}{\partial x^k}. $$
**Proof.** I follow [Courant:1937] and [Courant:1936], and in order to motivate the final argument, I first look at the case of dimension one. The fundamental theorem of calculus tells us that

\[ f(x) - f(0) = \int_0^x f'(s) \, ds. \]

An easy estimate tells us that the integral is \( O(x) \), but a simple trick will do better. If we set \( s = tx \) this equation becomes

\[ f(x) = f(0) + x \int_0^1 f'(tx) \, dt, \]

and the integral

\[ f_i(x) = \int_0^1 f'(tx) \, dt \]

is a smooth function of \( x \). Induction gives us

\[ f(x) = \sum_{k=m}^n c_k x^k + x^m f_m(x) \]

with \( f_m(x) \) smooth. An easy calculation tells us that \( c_k = f^{(k)}(0)/k! \).

In any dimension, for \( x \) in \( U \) define

\[ \varphi(t) = f(tx). \]

Then

\[ \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t) \, dt, \]

which because of our assumption and the chain rule translates to

\[ f(x) - f(0) = \sum_{i=1}^n x_i \int_0^1 \frac{\partial f}{\partial x_i}(tx) \, dt = \sum_{i=1}^n x_i f_i(x) \]

with \( f_i(x) \) smooth. Again apply induction.

Let \( m_p \) be the maximal ideal of \( C^\infty(M) \) consisting of functions on \( M \) vanishing at \( p \). As a consequence of the previous Lemma:

**1.3. Proposition.** Suppose \( U \) to be a convex open neighbourhood of the origin in \( \mathbb{R}^n \). The ideal \( m_p^U \) in \( C^\infty(U) \) is the space of all smooth functions \( f \) on \( U \) all of whose derivatives \( \partial^k f / \partial x^k \) with \( |k| < m \) vanish at \( p \).

Very roughly, a **vector bundle** \( B \) over \( M \) is a smooth manifold equipped with a projection onto \( M \), such that locally \( B \) and the projection possess a product structure—i.e. over small open sets \( U \) the bundle looks like \( U \times \mathbb{R}^n \). Its space of sections over \( U \) is therefore isomorphic to a sum of \( n \) copies of \( C^\infty(U) \). In this essay, I use this as the definition:

**Definition.** A vector bundle is a sheaf \( \Gamma(U) \) on \( M \) which is a module over the sheaf of smooth functions on \( M \), locally free in the sense that in some neighbourhood \( U \) of every \( p \) the module \( \Gamma(U) \) is isomorphic to a sum of \( n \) copies of \( C^\infty(U) \).

The **fibre** of the bundle over a point \( p \) is the quotient of \( \Gamma(M) \) by the submodule \( m_p \Gamma(M) \). For \( \gamma \) in \( \Gamma(U) \), let \( \gamma(p) \) be its image in the fibre over \( p \).

**How does one recover the geometric bundle itself from the module defining it?** Define the geometric bundle to be the set of all pairs \((p, v)\) with \( p \) in \( M \) and \( v \) in the fibre over \( p \). How to make this into a smooth manifold? We only need to do it locally on \( M \). Suppose \( U \) to be a neighbourhood of \( p \) with \( \Gamma(U) \) isomorphic to the sum of \( n \) copies of \( C^\infty(U) \). Choosing a basis \((s_i)\) of \( \Gamma(U) \) as a module over \( C^\infty(U) \) identifies the geometric bundle with \( U \times \mathbb{R}^n \).

The advantage of defining a bundle by its sections is that it is frequently not obvious how to define the geometric object, but relatively simple to specify its sections. We’ll see some examples in the next section. I’ll usually leave it as an exercise to verify local freeness.
The point of vector bundles is that they are not globally trivial. For example, define $M$ to be $\mathbb{P}^1(\mathbb{R}^2)$, which is canonically defined to be the set of lines through the origin in $\mathbb{R}^2$. It is the union of two open sets $U_0$ and $U_\infty$, each isomorphic to $\mathbb{R}$. The first is the set of lines through $(0, 1)$, the second the set of lines through $(1, 0)$.

I define a vector bundle $B$ over this to be the set of pairs $(\ell, x)$ where $\ell$ is a line through the origin in $\mathbb{R}^2$ and $x$ is a point on $\ell$. The fibration takes $(\ell, x)$ to $\ell$, and each fibre is isomorphic, although not canonically, to $\mathbb{R}$. Over either $U = U_i$, the space of sections is isomorphic to $C^\infty(U)$, but the space of all sections is not a free module over $C^\infty(M)$ because of:

1.4. Proposition. Every global section of $B$ over $M$ vanishes somewhere.

Of course this wouldn’t be true if the space of global sections were free over $C^\infty(M)$.

Proof. The manifold $M$ is the quotient of the unit circle $S$ in $\mathbb{R}^2$ by $\pm 1$, since every line through the origin in $\mathbb{R}^2$ passes through opposite points on the unit circle. If $x(t)$ is a non-zero point on $\ell$ then $x/\|x\|$ is a point on $S$, so a non-zero section of $B$ would give rise to a map back from $M$ to $S$. However, there are no such maps, as one can verify by considering what such a map has to look like on each $U_i$. (Traveling completely around the world you lose or gain one day, not two.)

In other words, this vector bundle is essentially a Möbius strip.

2. Vector fields and differential operators

A vector $V$ at a point $p$ of $\mathbb{R}^n$ is a relative displacement from that point—the displacement of $p$ to $p + V$. We can add such vectors and multiply them by a scalar. But this only makes sense because of the linear structure of $\mathbb{R}^n$. If we impose a coordinate system in the neighbourhood of $p$ which is non-linear with respect to the original one, this definition is no longer valid. What is the intrinsic definition of a vector, independent of a choice of a coordinate system?

It is a velocity, or rate of relative displacement. If $p$ is a point of $\mathbb{R}^n$ and $V$ a vector, one can define the directional derivative determined by $V$. Given a smooth function $f$ defined in the neighbourhood of $p$, restrict it to the path $\gamma: t \mapsto p + tv$ running through $p$ in the direction of $V = (v_i)$, and take its derivative at $t = 0$:

$$V \cdot f = \frac{d}{dt} f(p + tv) \bigg|_{t=0} = \sum_{i=1}^n v_i \frac{\partial f}{\partial x_i}(p).$$

This still seems to depend on the linear structure of a coordinate system, but in fact it does not. Suppose $\gamma(t) = (x_i(t))$ to be any smooth path through $p$ with $\gamma(0) = p$. Its velocity at $t = 0$ is the vector $\gamma'(0) = (dx_i/dt)$ evaluated at $t = 0$. By the chain rule we have

$$\frac{df}{dt} = \sum_i \frac{\partial f}{\partial x_i} \frac{dx_i}{dt},$$

so that as $\gamma$ is concerned its value at 0 depends only on the velocity $\gamma'(0)$. Therefore any path at all through $p$ with the same velocity vector as the path $\gamma(t) = p + tv$ will give the same directional derivative as $V$.

If all partial derivatives $\partial f/\partial x_i$ vanish then at $p$ of course the directional derivative $V \cdot f$ does, too. According to Proposition 1.3, the directional derivative $f \mapsto V \cdot f$ at $p$ induces a linear map from $\mathcal{O}_p/m_p^2$ to $\mathbb{R}$ that vanishes on the constants. Such maps are determined completely by their restrictions to $m_p/m_p^2$, so that tangent vectors at $p$ may be canonically defined as elements of the linear dual of $m_p/m_p^2$.

A smooth vector field on an open set $U$ of $\mathbb{R}^n$ assigns to every point $x$ of $U$ a vector $(v_i(x))$, where the $v_i$ are smooth functions on $U$. Applying directional derivatives at each point, it becomes a differential operator taking a smooth function $f$ on $U$ to the function

$$[V \cdot f](x) = \sum v_i(x) \frac{\partial f}{\partial x_i}(x).$$
2.1. Proposition. Every smooth vector field on an open subset $U$ of $\mathbb{R}^n$ induces a $\mathbb{C}$-linear derivation of $C^\infty(U)$. Conversely, any $\mathbb{C}$-linear derivation of $C^\infty(U)$ is determined by a smooth vector field.

I recall that a derivation of a ring $R$ over $\mathbb{C}$ is a $\mathbb{C}$-linear map $D$ from $R$ to itself such that

$$D(rs) = D(r)s + rD(s)$$

for all $r, s$ in $R$. The condition of $\mathbb{C}$-linearity means neither more nor less than that it annihilates the constants.

Proof. One way is simple—Leibniz’ rule tells us that a vector field is a derivation.

As for the converse, suppose that $D = \sum v_i \frac{\partial}{\partial x_i}$ is a vector field, where each $v_i$ is a smooth function. Then the coefficient $v_i$ may be recovered as $D(x_i)$. So, suppose now that $D$ is a derivation of $C^\infty(U)$. Let $v_i = D(x_i)$. We want to show that

$$D(f) = \sum_i v_i \frac{\partial f_i}{\partial x_i},$$

and it suffices to show this at any point $p = (p_i)$ of $U$. In some convex neighbourhood of $p$

$$f(x) - f(p) = \sum (x_i - p_i)f_i(x)$$

$$Df = \sum D(x_i)f_i + (x_i - p_i)Df_i$$

$$[Df](p) = \sum v_i(p)f_i(p)$$

$$[Df](p) = \sum v_i(p) \frac{\partial f_i}{\partial x_i}(p).$$

I now make the

Definition. A vector field on an arbitrary manifold $M$ is a derivation of $C^\infty(M)$.

We have just seen that this agrees with the definition for $M$ an open subset of $\mathbb{R}^n$. We shall next see that a vector field in the sense just defined induces a vector field on any open subset of $M$, so that the new definition agrees completely with the more obvious one. Here is the first step towards that:

2.2. Lemma. A vector field $V$ on $M$ is local in the sense that the restriction of $V \cdot f$ to an open subset $U$ of $M$ depends only on the restriction of $f$ to $U$.

Another way to say this is that the support of $V \cdot f$ is contained in the support of $f$.

Proof. If $f$ and $f_*$ have the same restriction to $U$, then $f - f_*$ vanishes on $U$. Thus it must be shown that if $f(x) = 0$ throughout any small disk in a coordinate neighbourhood $U$ of $M$ then $Df$ vanishes throughout $U$ as well. Suppose $x$ to be a point of $U$. We can find a smooth function $\varphi$ which is 1 outside $U$ and vanishes at $p$. Then $\varphi f = f$ and

$$[D(\varphi f)](p) = [Df](p)\varphi(p) + f(p)[D\varphi](p) = 0.$$

So $Df$ vanishes at every point of $U$.

2.3. Lemma. If $M$ is an arbitrary manifold, $U$ an open subset of $M$, and $D$ a vector field on $M$, there exists a unique derivation of $C^\infty(U)$ that agrees with the restriction of functions from $M$ to $U$.

In other words, sections of vector fields form a sheaf.
Proof. Suppose \( f \) in \( C^\infty(U) \). We must define \( Df \) as a smooth function on \( U \). If we are given a covering \( \{ U_\alpha \} \) of \( U \) by open sets, then we have a sequence of restriction maps

\[
0 \longrightarrow C^\infty(U) \longrightarrow \prod_\alpha C^\infty(U_\alpha) \longrightarrow \prod_{\beta \gamma} C^\infty(U_\beta \cap U_\gamma).
\]

Since \( U \) may be covered by disks in coordinate patches, it suffices to define \( Df \) on such a disk, and show compatibility on two overlapping disks.

Let \( U_* \) be a disk whose closure is contained in a coordinate patch, \( f \) in \( C^\infty(U_*) \). We can find a smooth function \( f_* \) on all of \( M \) restricting to \( f \) on \( U_* \). Define \( Df \) in \( U_* \) to be the restriction of \( Df_* \) to \( U_* \). If \( U_{**} \) is any open subset of \( U_* \), the restriction of \( Df \) to \( U_{**} \) only depends on the restriction of \( Df_* \) to \( U_{**} \), assuring compatibility.

If \( X \) and \( Y \) are two vector fields, their Poisson bracket \([X, Y]\) is the difference \( XY - YX\). It is a priori a differential operator possibly of order 2, but in fact it is a vector field. This can be seen by a local computation, but also by checking directly that it is a derivation.

If \( F \) is a smooth map from one manifold to another and \( f \) is a smooth function on the target manifold then the composite \( F^*f(x) = f(F(x)) \) is a smooth function on the source. Thus \( F \) gives rise to a map from each \( \mathcal{C}_F(p) \) to \( \mathcal{C}_p \), hence to a dual map from vectors at \( p \) to vectors at \( F(p) \). In particular, if \( \gamma(t) \) is a path from a segment \( J \) of \( \mathbb{R} \) to a manifold, each \( t \) in \( I \) determines a vector \( \gamma'(t) \) at \( \gamma(t) \). Diffeomorphisms similarly transport vector fields.

The vector bundle associated to the sheaf of derivations is the tangent bundle \( T_M \). Its dual bundle is the cotangent bundle \( T^*_M \), whose fibre at a point \( p \) is \( m_p/m_p^m \). It is not quite obvious how to directly construct that bundle, nor how to construct more generally a bundle over \( M \) whose fibre at \( p \) is \( m_p^{m-1}/m_p^m \), nor yet a bundle whose fibre at \( p \) is \( \mathcal{C}_p/m_p^m \).

I don’t know to whom the following trick is due. Consider in \( M \times M \) the diagonal \( \Delta = \Delta_M \), and the ideal \( I = I_\Delta \) of functions vanishing on \( \Delta \). For each open set \( U \) in the first factor \( M \) consider the corresponding subset of \( \Delta \). Define \( \Gamma(U) \) to be the direct limit of the quotients \( C^\infty(U_\alpha)/I^m C^\infty(U_\alpha) \) as \( U_\alpha \) varies over the open sets of \( M \times M \) with \( U_* \cap \Delta = U \). Projection from \( \Delta \) onto the first factor makes this a module over \( C^\infty(U) \). It turns out that this defines a locally free sheaf \( \mathcal{C}^m \) on \( M \) whose fibre at \( p \) may be identified with \( \mathcal{C}_p/m_p^m \). For example, if \( M = \mathbb{R} \), one can verify this by changing variables from \( (x, y) \) to \( (x, y - x) \) along the diagonal. Similarly, sections of the bundle with fibres \( m/m^m \) can be identified with \( I/I^m \) as a module over \( C^\infty(U) \).

A smooth differential operator on a manifold \( M \) is an operator \( D \) on \( C^\infty(M) \) locally expressed as a finite sum over multi-indices \( k \):

\[
D = \sum f_k(x) \frac{\partial^k}{\partial x^k}
\]

where the \( f_k \) are smooth functions. The order of \( D \) is the largest \( m \) for which some \( f_k \neq 0 \) with \( |k| = \sum k_i = m \). The ring of differential operators on \( M \) is generated by \( C^\infty(M) \) and smooth vector fields. The differential operators define a sheaf which is canonically the dual of \( \mathcal{C}^m \). I leave as exercise:

2.4. Proposition. The quotient \( D^m/D^{m-1} \) is isomorphic to the smooth homogeneous functions of degree \( m \) on the cotangent bundle, which may be identified with the symmetric product of the tangent bundle \( S^m(T_M) \).

Any vector field on a smooth manifold determines on it a flow, which determines a trajectory through each point tangential to the vector field at that point. In local coordinates, this comes down to solving an ordinary differential equation. More precisely, if the vector field is \( X \) then the trajectory of the flow starting at the point \( x \) of \( G \) is a smooth path \( t \mapsto E(t) = E_X(t) \) satisfying the differential equation \( E'(t) = X E(t) \) and initial condition \( E(0) = x \). In local coordinates, if \( X = \sum X_i(x) \partial/\partial x_i \) and \( E(t) = (x_i(t)) \) then \( d/dt \) maps to \( \sum_i (dx_i/dt)(\partial/\partial x_i) \) and the differential equation is that determined by system of equations

\[
dx_i/dt = X_i(x)\,.
\]
Remark. Tangent bundles are generally non-trivial, and do not generally possess non-zero sections. For example, there are no vector fields on the unit sphere in $\mathbb{R}^3$ that do not vanish anywhere. This can be proved by considering what those fields must be like on each of the north and south hemispheres and then comparing what you get along the equator, their common boundary.

3. The Lie algebra of a Lie group

Now suppose $G$ to be a Lie group—that is to say, a smooth differentiable manifold with a smooth structure as topological group. This means that multiplication $G \times G \to G$ and the inverse $G \to G$ are smooth maps.

The Lie algebra $\text{Lie}(G)$ of $G$ is defined to be the space of left-invariant vector fields on $G$. If $g$ is any element of $G$ and $f$ a smooth function defined in the neighbourhood $U$ of $g$ then $L_g^* f$ is the pullback of $f$ to $g^{-1} U$, a neighbourhood of $I$, defined by the formula $L_g^* f(x) = f(gx)$. This induces a linear map $L_{g,*}$ from the tangent space at $I$ to that at $g$:

$$\langle L_{g,*} X, f \rangle = \langle X, L_g^* f \rangle$$

That the function of $g$ is smooth is a consequence of the following elementary remark: suppose $M$ and $N$ to be manifolds, $n$ a point of $N$ and $V$ a vector at $n$. It determines at each point of $M \times \{n\}$ a ‘vertical’ vector field. For any smooth function on $M \times N$ the function $V \cdot f$ evaluated on $M$ is smooth. This is to be applied to $G \times \{1\}$. Conversely, any element $X$ of $\mathfrak{g}$ determines a smooth vector field, which I’ll call $R_X$ for reasons to be seen in a moment, on all of $G$. It is invariant with respect to all left $G$-translations. It is unique subject to the two conditions that it be left-invariant and agree with $X$ at $I$. Therefore:

3.1. Proposition. The map taking a left-invariant vector field to its value at 1 is an isomorphism of $\text{Lie}(G)$ with the tangent space at 1.

The flow associated to $R_X$ is also left-invariant—if $E_X(t)$ is the trajectory starting at $I$, the trajectory at $g$ is $g E_X(t)$. The flow on $G$ is generated by the left translations of the trajectory $E_X(t)$ starting at $I$. The defining equation for $E_X$ implies that $E_X(s) E_X(t) = E_X(s + t)$, which because of continuity implies that $E_X(t) = E_{tX}(1)$. This last is written as the Lie exponential map $\exp(tX)$. That is to say

$$R_X F(g) = \frac{d}{dt} F\left(g \exp(tX)\right) \bigg|_{t=0}.$$ 

Let’s look at some examples.

Example. The additive group $\mathbb{R}$ of real numbers. The space $\mathfrak{g}$ is $\mathbb{R}$. The invariant vector fields are the $a \frac{d}{dx}$ where $a$ is a constant. The corresponding flow $x(t)$ satisfies the differential equation

$$x'(t) = a$$

and the flow starting at $x$ is $x(t) = x + at$.

Example. The multiplicative group of non-zero real numbers $\mathbb{R}^\times$. The space $\mathfrak{g}$ may again be identified with $\mathbb{R}$, where $a$ corresponds to the vector $a d/dx$ at 1. Multiplicative translation by $x$ takes $a$ at 1 to $ax \frac{d}{dx}$. The invariant vector fields are therefore the $ax \frac{d}{dx}$, the differential equation of the flow $x'(t) = ax(t)$, and the corresponding flow at 1 takes $t$ to $\exp(ta)$. The connected component of positive real numbers is isomorphic to the additive group through the exponential map, but both $\exp$ and its inverse $\log$ are analytic functions. In terms of algebraic structure the two groups are thus distinct.

Example. The multiplicative group $\mathbb{S}$ of complex numbers of unit magnitude. The space $\mathfrak{g}$ is again $\mathbb{R}$, with $a$ corresponding to $a d/d\theta$ at 1, where $\theta$ is the argument. The invariant vector fields are the $a d/d\theta$. The flow starting at 1 is the complex exponential function $\exp iat$.

Example. The multiplicative group of non-zero complex numbers $\mathbb{C}^\times$. The space $\mathfrak{g}$ may be identified with $\mathbb{C}$. The flows at 1 are the complex exponential $\exp(tz)$, and the trajectories will usually be spirals. This group contains the previous two as subgroups.
Example. The multiplicative group $GL_n(C)$ of invertible real $n \times n$ matrices. This group is an open subset of the matrices $M_n(C)$, and the space $g$ may be identified with it. The invariant vector fields are the $X \partial/\partial X$ for invertible $X$. The flow at $I$ tangent to $A$ is the matrix exponential $\exp(tA)$, where

$$e^X = \exp(X) = I + X + \frac{X^2}{2!} + \cdots$$

The series converges for all $X$, and by the implicit function theorem maps a neighbourhood of $0$ in $M_n(C)$ isomorphically onto a neighbourhood of $I$ in $GL_n(C)$. We can calculate explicitly some simple examples for $n=2$:

$$\exp\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = \begin{pmatrix} e^a & 0 \\ 0 & e^b \end{pmatrix}$$

$$\exp\begin{pmatrix} 0 & x \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$$

$$\exp\begin{pmatrix} 0 & -t \\ t & 0 \end{pmatrix} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$ 

For matrices near enough to $I$ we can write out explicitly an inverse to the exponential map, by means of the series

$$\log(I + X) = X - \frac{X^2}{2} + \frac{X^3}{3} - \cdots,$$

which converges for $X$ small.

In general, the relationship between the group and the Lie algebra $g$ is intuitive. Elements in a group very close to $I$ are of the form $I + X + \text{higher order terms}$, where $X$ is a very small element of the Lie algebra. Intuitively, the Lie algebra represents very, very small motions in the group, those not differing much from the identity. The exponential map makes this rough idea precise.

3.2. Proposition. On any Lie group, the exponential map has these properties:

(a) if $X$ and $Y$ commute then $e^{X+Y} = e^X e^Y$;

(b) the inverse of $e^X$ is $e^{-X}$;

(c) if $g$ is invertible then $e^{\text{Ad}(g)} = g e^X g^{-1}$;

(d) the exponential of a diagonal matrix with entries $d_i$ is the diagonal matrix with entries $e^{d_i}$;

(e) the exponential of the transpose is the transpose of the exponential;

(f) the determinant of $e^X$ is $e^{\text{trace}(X)}$.

Proof. All are straightforward except possibly (f), for which I’ll give two proofs.

First proof. If $X$ is an arbitrary complex matrix then some conjugate $X_* = YXY^{-1}$ will be in Jordan form $S + N$, where $S$ is diagonal, $N$ is nilpotent, and the two commute. The matrices $X$ and $X_*$ will have the same determinant and the same trace, so by (c) we may as well let $X = X_*$. But

$$e^X = e^S e^N$$

will again be in Jordan form, with diagonal equal to that of $e^S$. Hence the trace of $Y$ is that of $S$ and the determinant of $e^S$ is that of $e^S$. But for $S$ the claim is clear.

Second proof. It suffices to show that the function $D(t) = \det(e^{tX})$ satisfies the ordinary differential equation

$$D'(t) = \text{trace}(X) D(t),$$

where $e^X = \exp(X) = I + X + \frac{X^2}{2!} + \cdots$.
Introduction to Lie algebras

since it certainly satisfies the initial condition \( D(0) = 1 \). But

\[
D'(t) = \lim_{h \to 0} \frac{D(t + h) - D(t)}{h} = \frac{\det (e^{(t+h)X}) - \det (e^{tX})}{h} = \frac{\det (e^{tX}) \det (e^{hX}) - \det (e^{tX})}{h} = \det (e^{tX}) \left[ \frac{\det (e^{hX}) - 1}{h} \right]
\]

which has as limit \( D(t) \) \( \text{trace}(X) \) since \( \det(I + hX + \cdots) = 1 + h \text{trace}(X) + \cdots \).

As one consequence, the Lie algebra of \( SL_n \) is the space of matrices with vanishing trace. This technique can be formalized. If \( G \) is a group defined by polynomials \( P(x) = 0 \) in \( M_n(C) \), then the tangent space at \( I \) is defined by conditions \( P(I + \varepsilon X) = 0 \) modulo second order terms in \( \varepsilon \). But \( P(I + \varepsilon X) = P(I) + \varepsilon(\nabla P, X) \) modulo \( \varepsilon^2 \), and this is just \( \varepsilon dP(X) \) since \( I \) is in the group and \( P(I) = 0 \). Therefore the Lie algebra is defined by the equations

\[
\langle dP(X), X \rangle = 0.
\]

This agrees with what we just showed for \( SL_n \), since the differential of \( \det \) is \( \text{trace} \). This will show, for example, that the Lie algebra of the symplectic group \( Sp \) is the space of matrices \( X \) such that \( X J + J X = 0 \) where

\[
J = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}
\]

Of course the exponential maps real matrices onto real matrices, so the tangent space of \( GL_n(R) \) at \( I \) may be identified with \( M_n(R) \).

Any tangent vector at \( I \) determines by right as well as left translation vector fields \( \Lambda_X \) as well as \( R_X \) on all of \( G \). Vector fields are differential operators, acting on scalar functions on a Lie group. The derivative of \( F \) at \( g \) with respect to \( R_X \) is

\[
R_X F(g) = \frac{d}{dt} F(g \exp(tX)) \Bigg|_{t=0},
\]

and that with respect to \( \Lambda_X \) is

\[
\Lambda_X F(g) = \frac{d}{dt} F(\exp(tX) g) \Bigg|_{t=0}.
\]

The vector fields \( \Lambda_X \) are those which are right-invariant.

If \( U \) and \( V \) are left-invariant so is \( UV - VU \), so we get from the Poisson bracket in this way the Lie bracket operation on a Lie algebra.

If \( G = GL_n \), we can ask for an explicit formula for \( [R_X, R_Y] \), where \( X \) and \( Y \) are given matrices. Let \( M \) be the tautological map from \( GL_n(R) \) to \( M_n(R) \), taking \( g \) to its matrix \( M(g) \). Thus

\[
R_X M(g) = \left[ \frac{d}{dt} M(g \exp(tX)) \right]_{t=0} = \left[ \frac{d}{dt} M(g \exp(tX)) \right]_{t=0} = M(g)X.
\]

Then

\[
\]
Therefore:

**3.3. Proposition.** In $M_n(\mathbb{R})$, the Lie algebra of $\text{GL}_n(\mathbb{R})$, the Lie bracket is the commutator:

$$[X, Y] = XY - YX.$$ 

There is one peculiarity about the right-invariant vector fields. If $\Lambda_X$ and $\Lambda_Y$ are two of these, then their Poisson bracket $[\Lambda_X, \Lambda_Y]$ is equal to $\Lambda_{[Y,X]}$ rather than $\Lambda_{[X,Y]}$. This is related to the following point—the left-invariant vector fields $R_X$ derive from the right-regular action of the group $G$ on smooth functions on $G$, according to which

$$[R_g F](x) = F(xg).$$

The right action on the space defines a left action of $G$ on functions, since

$$[R_g R_h F](x) = [R_h F](xg) = F(xgh) = [R_{gh} F](x), \quad R_{gh} F = R_g R_h F.$$ 

The left-regular action is defined by

$$L_g F(x) = F(g^{-1}x).$$

The associated left action of the Lie algebra is $L_X = \Lambda_{-X}$. Thus the bracket equation for right-invariant vector fields is, as it should be,

$$[L_X, L_Y] = L_{[X,Y]}.$$

The adjoint action has a simple expression in the case of $\text{GL}_n$—it turns out to be the obvious matrix calculation:

**3.4. Proposition.** For any $g$ in $\text{GL}_n(\mathbb{R})$ and $X$ in $M_n(g)$

$$\text{Ad}(g)X = gXg^{-1}.$$ 

**Proof.** Very similar to that of Proposition 3.3.

Thus we can see also why the earlier claim about $\text{ad}$ is true: $X$ in $\mathfrak{g}$ gets mapped to the endomorphism

$$\text{ad}_X: Y \mapsto [X,Y]$$

since

$$(I + tA + \cdots)X(I + tA + \cdots)^{-1} = (I + tA + \cdots)X(I - tA + \cdots)$$

$$= X + t(AX - XA) + \cdots.$$
4. Normal coordinates

The exponential map \( g \to G \) is a local diffeomorphism, hence a coordinate system on \( g \) induces a coordinate system in an open neighbourhood \( U \) of the identity on \( G \). If we choose a basis \((X_i)\) of \( g \), then the coordinate system maps \( \exp(t_1X_1 + \cdots + t_nX_n) \) to \((t_1, \ldots, t_n)\).

Let \( m \) be the maximal ideal of functions in \( C^\infty(U) \) vanishing at 1. Its power \( m^m \) is the subset of smooth functions on \( U \) such that \( \partial^k f / \partial x^k = 0 \) for all \(|k| < m\), or equivalently those which can be represented as a sum \( t^k f_k \) with \(|k| = m\) and each \( f_k \) in \( C^\infty(U) \). The vector field \( X_i \) and \( \partial / \partial t_i \) agree at 1, hence \( X_i = \partial / \partial t_i \sum_{j \neq i} f_j \partial / \partial x_j \) with each \( f_j \) in \( m \). More generally, as a consequence, if \(|k| \leq |\ell|\) then \( X^k t^\ell = \{ \ell! \) if \( k = \ell \) \( 0 \) otherwise.

Therefore:

4.1. Proposition. The pairing \( \langle X, f \rangle = X F(1) \) identifies \( U_{n-1}(g) \) with the annihilator of \( m^n \), and \( U_n(g)/U_{n-1}(g) \) is isomorphic to \( S^n(g) \).

4.2. Corollary. The operators \( R_X \) for \( X \) in \( U(g) \) are exactly the left-invariant differential operators on \( G \).

Proof. Because \( U(g) \) exhausts the invariant symbols.

5. The Lie algebra of \( SL(2) \)

Now let \( G = SL_2(\mathbb{R}) \) and \( g = sl_2 \) its Lie algebra, the space of \( 2 \times 2 \) matrices of trace 0. Let \( A \) be the subgroup of diagonal matrices, \( N \) that of unipotent upper triangular matrices, \( K \) the special orthogonal group \( SO_2 \). Several distinct bases for \( g \) are useful in different situations. One very useful basis is this:

\[
e_+ := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad h := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \kappa := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},
\]

which are, respectively, elements in \( n, a, k \) (the Lie algebras of \( N, A, \) and \( K \)). That \( g = n + a + k \) is the infinitesimal version of the Iwasawa decomposition \( G = NAK \).

Another useful basis is made up of \( e_+, h \) and

\[
e_- := \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}.
\]

I depart here from the standard convention, which replaces \( e_- \) by its negative. I follow [Tits:1966] in this. The advantage of the notation he and I use is symmetry: the map \( e_\pm \mapsto e_\mp, h \mapsto -h \) is the canonical involution of \( sl_2 \), an automorphism of order two.

This new element \( e_- \) spans the Lie algebra \( \mathfrak{p} \) of the lower triangular unipotent elements of \( SL_2 \), and the decomposition \( g = n + a + \mathfrak{p} \) is an infinitesimal version of the Bruhat decomposition \( G = PNP \cup Pw \) where

\[
w = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.
\]

This \( w \) is the same as \( \kappa \), but it is important to distinguish them—one lies in \( SL_2 \) and the other in \( sl_2 \).
We have the Lie algebra bracket formulae
\[ [h, e_{\pm}] = \pm 2 e_{\pm}, \quad [e_{+}, e_{-}] = -h. \]

Thus \( e_{\pm} \) are eigenvectors of \( \text{ad}_h \), called root vectors of \( \mathfrak{g} \) with respect to \( a \).

There is a third useful basis. The element \( \kappa \) also spans the Lie algebra of the torus \( K \) of \( G \), but it is a compact one, made up of rotation matrices. Inside \( G(\mathbb{C}) \) the two tori \( A \) and \( K \) are conjugate, and more precisely the Cayley transform
\[
C = \begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}
\]
conjugates \( K(\mathbb{C}) \) to \( A(\mathbb{C}) \). This can be seen geometrically—\( K \) is the isotropy subgroup of \( \text{SL}_2(\mathbb{R}) \) acting by Möbius transformations on the upper half plane, while \( A \) preserves the origin. Consistently with this, the Cayley transform takes \( i \) to \( 0 \). In the complexified Lie algebra
\[
\text{Ad}(C^{-1}) h = i\kappa.
\]

Since \( K_{\mathbb{C}} \) is a torus, the complexified Lie algebra decomposes into root spaces, the images under \( C \) of those for \( A_{\mathbb{C}} \). Explicitly they are spanned by
\[
x_{\pm} = \begin{bmatrix} 1 & \mp i \\ i & 1 \end{bmatrix}.
\]

We have now
\[
[k, x_{\pm}] = \pm 2ix_{\pm}
\]
and
\[
\text{Ad}(C^{-1}) e_{\pm} = (1/2) x_{\mp}.
\]

5.1. Lemma. The center of \( \mathfrak{sl}_2 \) is \( \{0\} \), and any element of \( \mathfrak{sl}_2 \) can be expressed as a linear combination of commutators \( [X, Y] \).

Proof. It suffices to verify it for elements of a basis, which I leave as an exercise.

Yet more strongly:

5.2. Proposition. Any homomorphism from \( \mathfrak{sl}_2 \) to another Lie algebra is either trivial or injective.

Proof. If it is not injective, the image must be a Lie algebra of dimension 1 or 2. If it is onto an abelian algebra, then \( [X, Y] \) is mapped to 0 for all \( X \) and \( Y \), and therefore by the Lemma so is all of \( \mathfrak{sl}_2 \). If the image is not abelian, then the image must be the unique non-abelian algebra of two dimensions described in Proposition 7.2, and in that case too, since that algebra has the trivial algebra as quotient, all of \( \mathfrak{sl}_2 \) is mapped to 0.

6. Vector fields associated to \( G \)-actions

The action of a Lie group \( G \) on a manifold \( M \) determines also vector fields corresponding to vectors in its Lie algebra, the flows along the orbits of one-parameter subgroups \( \exp(tX) \). Thus the vector at a point \( m \) corresponding to \( X \) in \( \mathfrak{g} \) is the image of \( d/dt \) at \( t = 0 \) under the map from \( \mathbb{R} \) to \( M \) taking \( t \) to \( \exp(tX) \cdot m \). This is a special case of the general problem of calculating the image of \( d/dt \) under a map onto a manifold. If we are given local coordinates around \( m \) and the map takes
\[
t \mapsto (x_i(t))
\]
then
\[
\frac{d}{dt} \mapsto \sum \frac{dx_i}{dt} \frac{\partial}{\partial x_i}.
\]


This might seem sometimes to involve a formidable calculation, and it is often useful to use Taylor series to simplify it. The point is that it is essentially a first order computation in which terms of second order can be neglected. Roughly speaking, up to first order \( \exp(\epsilon X) = I + \epsilon X \), so the element \( X \) in \( \mathfrak{g} \) determines at \( m \) the vector

\[
\frac{(I + \epsilon X) \cdot m - m}{\epsilon}
\]

where we may assume \( \epsilon^2 = 0 \). The coefficients of the \( \partial/\partial x_i \) are read off as the coordinates of \( \epsilon \) in the expression for \( (I + \epsilon X) \cdot m \).

Let’s look at the example of \( \text{SL}_2(\mathbb{R}) \) acting on the upper half plane \( \mathcal{H} \) by Möbius transformations

\[
z \mapsto az + b \quad \text{c}z + d.
\]

In the formulas for vector fields associated to elements of the Lie algebra \( \mathfrak{sl}_2 \), some simplification is possible because the vector fields are real and the group acts holomorphically. The natural result of these calculations will a complex-valued function. A complex analytic function \( f(z) = p + iq \) is to be interpreted as as a real vector field according to the formula

\[
p \frac{d}{dx} + q \frac{d}{dy} = (p + iq) \left( \frac{1}{2} \right) \left( \frac{d}{dx} - i \frac{d}{dy} \right) + (p - iq) \left( \frac{1}{2} \right) \left( \frac{d}{dx} + i \frac{d}{dy} \right) = f(z) \frac{d}{dz} + \overline{f(z)} \frac{d}{\overline{z}}.
\]

6.1. Proposition. We have

\[
\Lambda_{e^+} = \frac{\partial}{\partial x},
\]

\[
\Lambda_h = 2x \frac{\partial}{\partial x} + 2y \frac{\partial}{\partial y},
\]

\[
\Lambda_{e^-} = (x^2 - y^2) \frac{\partial}{\partial x} + 2xy \frac{\partial}{\partial y}.
\]

Proof. • The simplest is \( e^+ \). Here

\[
I + \epsilon e^+ = \begin{bmatrix} 1 & \epsilon \\ 0 & 1 \end{bmatrix}
\]

and this takes

\[
z \mapsto \frac{z + \epsilon}{1} = z + \epsilon, \quad (x, y) \mapsto (x + \epsilon, y)
\]

Therefore

\[
e^+ \sim \partial/\partial x.
\]

• Now for \( h \). Here

\[
I + \epsilon h = \begin{bmatrix} 1 + \epsilon & 0 \\ 0 & 1 - \epsilon \end{bmatrix}
\]

and this takes

\[
z \mapsto \frac{(1 + \epsilon)z}{(1 - \epsilon)} = z(1 + \epsilon)(1 + \epsilon + \epsilon^2 + \cdots) = z(1 + 2\epsilon) = z + 2\epsilon z,
\]

\[
(x, y) \mapsto (x + 2\epsilon x, y + 2\epsilon y)
\]

\[
(x, y) \mapsto (x + 2\epsilon x, y + 2\epsilon y)
\]
so
\[ h \mapsto 2x \frac{\partial}{\partial x} + 2y \frac{\partial}{\partial y} \]

\[ I + \varepsilon e_{-} = \begin{bmatrix} 1 & 0 \\ \varepsilon & -1 \end{bmatrix} \]
\[ z \mapsto \frac{z}{\varepsilon - z + 1} = z + \varepsilon z^2 \]
\[ e_{-} \mapsto (x^2 - y^2) \frac{\partial}{\partial x} + 2xy \frac{\partial}{\partial y} \]

Part II. Lie algebras on their own

7. Abstract Lie algebras

For \( X, Y, Z \) in the Lie algebra of a Lie group \( G \) we have an identity of differential operators
\[
[X, [Y, Z]] = X(YZ - ZY) - (YZ - ZY)X \\
= XYZ - XZY - YZX + ZYX \\
[Y, [Z, X]] = Y(ZX - XZ) - (ZX - XZ)Y \\
= YZX - YXZ - ZXY + XZY \\
[Z, [X, Y]] = Z(XY - YX) - (XY - YX)Z \\
= ZXY - ZYX - XYZ + YXZ
\]
and summing we get
\[
[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0.
\]
This last equation is called the Jacobi identity. It is easy to recall to mind, since it is the sum of three terms obtained by applying a cyclic permutation of \( X, Y, Z \) to \([X, [Y, Z]]\).

There are several ways to interpret it, aside from the formal calculation. Even the formal computation can be analyzed intelligently. It can be seen immediately that the expression
\[
[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]]
\]
changes sign if any two of the variable are swapped. Swaps generate all of \( S_3 \), and one can conclude that the sum must be a scalar multiple of the alternating sum
\[
A(X, Y, Z) = \sum_{\sigma \in S_3} \text{sgn}(\sigma) \sigma(X, Y, Z) = XYZ - XZY + YZX - YXZ + ZXY - ZYX.
\]
But a simple look at the coefficients of \([X, [Y, Z]]\) in the Jacobi sum makes it clear that its twelve terms amount to \( A(X, Y, Z) - A(X, Y, Z) = 0 \).

**Definition.** Suppose \( g \) to be an arbitrary finite-dimensional vector space over an arbitrary field \( F \), and suppose that it is given a bilinear map \( (x, y) \mapsto [x, y] \) from \( g \times g \) to \( g \) itself. This is said to define \( g \) as a Lie algebra if this bilinear map is anti-symmetric and satisfies Jacobi’s identity.

It is not immediately apparent that this definition captures completely what one wishes. One justification is the ‘Third Theorem of Lie’ (apparently first proved by Eli Cartan): If \( F = \mathbb{R} \) or \( \mathbb{C} \) then every Lie algebra is the Lie algebra of an analytic group over \( F \). A clear exposition can be found in LG §5.8 of [Serre:1965]. A related result, with a more direct argument, is Serre’s construction of ‘group chunks’ in LG §5.4.
Here is another way to understand the Jacobi identity. Let $\mathfrak{g}$ be an arbitrary Lie algebra defined over $\mathbb{R}$, and let $A = \text{Aut}(\mathfrak{g})$ be the group of automorphisms of $\mathfrak{g}$. This is an algebraic group defined over $\mathbb{R}$. It is by definition embedded in $\text{GL}_2(\mathfrak{g})$, and this representation gives rise to its differential, an embedding of its Lie algebra $\mathfrak{a}$ in the space of linear endomorphisms of $\mathfrak{g}$. For $a$ in $\mathfrak{a}$, $t$ in $\mathbb{R}$ let $\alpha_t$ be the automorphism $\exp(ta)$ of $\mathfrak{g}$. Then

$$d\alpha_t(x) = \frac{d}{dt} \alpha_t(x) \bigg|_{t=0}$$

Since each $\alpha_t$ is an automorphism of $\mathfrak{g}$, the product rule for derivatives implies that $d\alpha_t$ is a derivation:

$$d\alpha_t([x, y]) = [d\alpha_t(x), y] + [x, d\alpha_t(y)].$$

7.1. Proposition. The space of derivations of $\mathfrak{g}$ is the Lie algebra of $\text{Aut}(\mathfrak{g})$.

Proof. The group of derivations of $\mathfrak{g}$ is an algebraic subgroup of $\text{GL}(\mathfrak{g})$. We know from a remark in Part I that an element $X$ in $\mathfrak{gl}(\mathfrak{g})$ is in its Lie algebra if and only if $I + \varepsilon X$ lies in the group, where $\varepsilon^2 = 0$. This happens if and only if $X$ is a derivation.

One can also verify directly that if $X$ is a derivation then $\exp(X)$ is an automorphism.

Among the automorphisms of the Lie algebra of a group $G$ are its inner automorphisms, given by the adjoint action. For $G = \text{GL}_n(\mathbb{R})$ this is matrix conjugation:

$$\text{Ad}(g)X = gXg^{-1}.$$ 

The map $\text{Ad}(g)$ is an automorphism of the Lie algebra structure. The differential of $\text{Ad}$ is

$$\text{ad}: X \mapsto \text{the map taking } Y \text{ to } [X, Y]$$

since

$$(I + tX + \cdots)Y(I + tX + \cdots)^{-1} = (I + tX + \cdots)Y(I - tX + \cdots)$$
$$= Y + t(XY - YX) + \cdots.$$ 

Because each $\text{Ad}(g)$ is an automorphism of $\mathfrak{g}$, each $\text{ad}_X$ is a derivation:

$$\text{ad}_X[Y, Z] = [\text{ad}_X Y, Z] + [Y, \text{ad}_X Z]$$

which is precisely the Jacobi identity.

From now on, unless I specify otherwise, Lie algebras will be assumed to be defined over an arbitrary coefficient field $F$ of characteristic 0.

One Lie algebra is the matrix algebra $\text{End}_F(V)$ with $[X, Y] = XY - YX$. To distinguish this as a Lie algebra from this as a ring, I’ll write the Lie algebra as $\mathfrak{gl}(V)$. If $V = F^n$ the endomorphism ring will be $M_n(F)$ and the Lie algebra will be $\mathfrak{gl}_n(F)$.

One mildly useful exercise is to classify all Lie algebras of dimension 1 or 2. For dimension 1, the only possibility is $F$ with the trivial bracket.

7.2. Proposition. There are, up to isomorphism, exactly two Lie algebras of dimension 2, the trivial one where all brackets are 0, and that with basis $X, Y$ where $[X, Y] = Y$.

The first is the Lie algebra of the additive group $F^2$, or of the group of diagonal matrices in $\text{GL}_2(F)$, while the second is that of the group of matrices

$$\begin{pmatrix} a & x \\ 0 & 1 \end{pmatrix}.$$ 

This second algebra $\mathfrak{g}$ fits into an exact sequence of Lie algebras

$$0 \to F \to \mathfrak{g} \to F \to 0.$$
Proof. Let $X, Y$ be a basis, and suppose $[X,Y] = aX + bY$. Suppose one of the coefficients, say $b$, is not zero. Then

$$[cX, aX + bY] = c[aX + bY] = ab(cX + bY)$$

which is equal to $aX + bY$ if $c = b^{-1}$. So if $Y_* = aX + bY$ and $X_* = cX$ we then have

$$[X_*, Y_*] = Y_*.$$

8. Representations

A **Lie algebra homomorphism** is a linear map $\varphi: g \to h$ such that

$$[\varphi(X), \varphi(Y)] = \varphi([X,Y]).$$

If $g = \exp(tX), h = \exp(tY)$ then

$$ghg^{-1} = I + t^2[X,Y] + \text{higher order terms}.$$

and therefore a homomorphism $\varphi$ of Lie groups induces a homomorphism $d\varphi$ of Lie algebras, its differential.

An **ideal** in $g$ is a linear subspace $h$ such that $[X,Y]$ lies in $h$ for all $X$ in $g$ and $Y$ in $h$. The Lie bracket on $g$ thus induces the structure of a Lie algebra on the quotient space $g/h$. Conversely, the kernel of any homomorphism of Lie algebras is an ideal.

A **representation** of a Lie algebra on a vector space $V$ is a Lie algebra homomorphism $\varphi$ into $\text{gl}(V)$. This is a linear map into $\text{End}(V)$ such that $\varphi([X,Y]) = \varphi([X,Y])$. In this way, $V$ becomes a module over $g$.

If $U$ and $V$ are $g$-modules then so is their tensor product:

$$X(u \otimes v) = Xu \otimes v + u \otimes Xv$$

and the space $\text{Hom}_F(U, V)$:

$$[Xf](u) = X(f(u)) - f(Xv).$$

In particular the linear dual $\hat{U}$ becomes the **dual** representation. So does the tensor algebra $\otimes^* V \otimes \otimes^* \hat{V}$, and the exterior algebra $\Lambda^* V$ embedded in it.

These definitions are compatible with, and indeed motivated by, the corresponding definitions of representations of a group, since

$$(I + tX + \cdots)(u \otimes v) = (I + tX + \cdots)u \otimes (I + tX + \cdots)v$$

$$= u \otimes v + tXu \otimes v + u \otimes tXv + \cdots.$$

Every Lie algebra has at least one representation, the adjoint representation by the linear maps $ad_X$ (in which it is Jacobi’s identity that verifies it is a representation).

A one-dimensional representation of a Lie algebra is called a **character**. If $\varphi$ is a character of $g$ then $\varphi([x,y]) = 0$. Define $Dg$ to be the span of all comutators $[x,y]$.

8.1. **Proposition.** The subspace $Dg$ is an ideal of $g$.

**Proof.** Immediate from the Jacobi identity.

Any character $\varphi$ vanishes on $Dg$. In fact:

8.2. **Proposition.** The characters of a Lie algebra may be identified with the linear dual of $g/Dg$.

In particular, if $g = Dg$ then the only character of $g$ is the trivial one. This is true, for example, of $\mathfrak{sl}_2$, according to Lemma 5.1.
9. Nilpotent Lie algebras

The simplest Lie algebras are the abelian ones, for which \([X, Y]\) identically vanishes. Next simplest are the nilpotent ones.

The upper central series of the Lie algebra \(g\) is the succession of ideals \(C^i(g)\) in \(g\) defined recursively:

\[
C^0 = g \\
C^{n+1} = [g, C^n].
\]

This is a weakly decreasing sequence. Each \(C^n\) is an ideal of \(g\), and each quotient \(C^n/C^{n+1}\) is in the centre of \(g/C^{n+1}\).

If \(g\) is any Lie algebra, let \(C(g)\) be its center. The lower central series of \(g\) is the succession of ideals \(C_i(g)\) in \(g\) defined recursively:

\[
C_0 = \{0\} \\
C_{n+1} = \text{the inverse image in } g \text{ of } C(g/C_n).
\]

Thus in particular \(C_1 = C(g)\). This is a weakly increasing sequence.

It can happen that both series are trivial, with \(C^0 = g\) and \(C_0 = 0\). This happens for \(sl_2\). The following result is about the other extreme:

**9.1. Proposition.** Suppose \(g\) to be a Lie algebra. The following are equivalent:

(a) some \(C^n = \{0\}\);  
(b) some \(C_n = g\);  
(c) the Lie algebra \(g\) possesses a strictly increasing filtration by ideals \(g_n\) such that \([g, g_{n+1}] \subset g_n\);  
(d) the Lie algebra \(g\) possesses a strictly increasing filtration by ideals \(g_n\) such that \([g, g_{n+1}] \subset g_n\) with each \(g_{n+1}/g_n\) of dimension one.

I leave this as an exercise.

The Lie algebra \(g\) is defined to be nilpotent if one of these conditions holds. Any quotient or subalgebra of a nilpotent Lie algebra is nilpotent.

One nilpotent Lie algebra is \(n_n\), the subalgebra of \(gl_n\) of matrices whose entries are 0 on and below the diagonal. It has as basis the matrices \(e_{i,j}\) with \(i < j\), with a single entry 1 at position \((i, j)\). We have

\[
e_{i,j}e_{k,l} = \begin{cases} 
  e_{i,l} & \text{if } j = k \\
  -e_{k,j} & \text{if } l = i \\
  0 & \text{otherwise.}
\end{cases}
\]

Hence \(C^d\) is the subspace spanned by the \(e_{i,j}\) with \(j - i > d\). For example, when \(n = 3\) we have the basis

\[
e_{1,2} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad e_{1,3} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad e_{2,3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}
\]

with

\[
[e_{1,2}, e_{2,3}] = e_{1,3}, \quad [e_{1,2}, e_{1,3}] = 0, \quad [e_{2,3}, e_{1,3}] = 0.
\]

The next theorem (which was apparently first proven by Killing, in spite of the name by which it is frequently called) asserts that the \(n_n\) are in a very strong sense universal. It is an immediate consequence of the definition that a Lie algebra \(g\) is nilpotent if and only if its image in \(gl(g)\) with respect to the adjoint map is contained in a conjugate of the nilpotent upper triangular matrices. A much weaker criterion is in fact valid.
An element of a Lie algebra is said to be nilpotent if the endomorphism $ad_n(X)$ is nilpotent or, equivalently, conjugate to an upper triangular nilpotent matrix. Every element $X$ of a nilpotent Lie algebra is nilpotent in this sense. Conversely:

9.2. Theorem. (Engel’s Criterion) A Lie algebra is nilpotent if and only if every one of its elements is nilpotent.

In other words, it is globally nilpotent if and only if it is locally nilpotent. This is immediate from:

9.3. Proposition. Suppose $\mathfrak{g}$ to be a Lie subalgebra of $\mathfrak{gl}_n$, and suppose that every element of $\mathfrak{g}$ is nilpotent in $M_n$. Then $\mathfrak{g}$ is contained in a conjugate of $\mathfrak{n}_n$.

Here is a modest reformulation, apparently first found in [Mermin:1994], that makes possible a constructive proof on which one can base a practical algorithm.

9.4. Lemma. Suppose $\mathfrak{g}$ to be a Lie subalgebra of $\mathfrak{gl}_n$. Either it is contained in a conjugate of $\mathfrak{n}_n$, or there exists $X$ in $\mathfrak{g}$ that is non-nilpotent in $M_n$.

By an easy induction argument, Lemma 9.4 follows from:

9.5. Lemma. Suppose $\mathfrak{g}$ to be a Lie subalgebra of $\mathfrak{gl}_n$. Either there exists $v \neq 0$ in $V$ with $Xv = 0$ for all $X$ in $\mathfrak{g}$, or there exists $X$ in $\mathfrak{g}$ which is not nilpotent in $M_n$.

The proof will in fact be constructive, and it would be easy to design a practical algorithm based on it. In this algorithm, we would start with a basis of matrices $X_i$ in $\mathfrak{g}$, and the output would be either (1) vector $v \neq 0$ with $X_i v = 0$ for all $i$, or (2) some linear combination of the $X_i$ that is certifiably not nilpotent.

Proof. I begin the proof of the Proposition by looking in detail at the case in which $\mathfrak{g}$ has dimension one—i.e. at a single linear transformation of a finite-dimensional vector space $V$.

9.6. Lemma. (Fitting’s Lemma) If $T$ is a linear transformation of the finite-dimensional vector space $V$, then $V$ has a unique $T$-stable decomposition

$$V = V(T^n) \oplus T^n V$$

for $n \geq \dim V$. The operator $T$ is nilpotent on $V(T^n)$ and invertible on $T^n V$.

Here $V(A)$ means the kernel of $A$.

Proof. I recall first of all a variant of Gauss-Jordan elimination. Given any finite set of vectors in a vector space $V$, let $M$ be a matrix with these as columns. Applying elementary column operations will reduce $M$ to a matrix in echelon form whose columns are a basis for the subspace of $V$ they span. This reduced matrix is in fact unique, and there is a bijection between subspaces and matrices in column echelon form. If $M$ is the matrix of a linear transformation, this allows us to compute its image. Working with row operations instead, one can compute its kernel.

Now suppose $T$ to be a linear transformation of $V$ and compute successively the images of the powers $T^n$. If $T^{n+1} V = T^n V$ then $T^n V = T^{m} V$ for all $m \geq n$. Therefore the weakly decreasing sequence $T^n V$ is eventually stable, and for some $n \leq \dim V$. Similarly, the kernels $V(T^n)$ of $T^n$ in $V$ are a weakly increasing sequence, also eventually stable. I write $\langle T \rangle V$ for $T^n V$ and $V \langle T \rangle$ for $\ker(T^n)$ with $n \geq 0$. If $TT^n V = T^n V$, the transformation $T$ is invertible on $T^n V$, and hence $\langle T \rangle V$ is complementary to $V \langle T \rangle$. The operator $T$ is nilpotent if and only if $V \langle T \rangle = V$ and $\langle T \rangle V = 0$.

Fitting’s Lemma asserts that every $v$ can be expressed as $v_0 + v_1$ with $v_0 \in V(T^n)$, $v_1 \in T^n V$. We can make this decomposition more explicit. Suppose the characteristic polynomial of $T$ factors as $Q(x) x^n$ with $Q(x)$ not divisible by $x$. Then $Q(T) = 0$ on $T^n V$. The Euclidean algorithm gives us $A(x)$, $B(x)$ such that

$$1 = A(x) x^n + B(x) Q(x) .$$

But then $A(T) T^n$ is the projection of $V$ onto $T^n V$, and $B(T) Q(T)$ is the projection onto $V(T^n)$.

The subspaces $V \langle T \rangle$ and $\langle T \rangle V$ are called the Fitting components of $T$. 
Now back to the proof of Lemma 9.5. Suppose \( g \) to be a Lie subalgebra of \( \mathfrak{gl}(V) \).

First an elementary observation: *If \( X \) is a nilpotent operator on \( V \), then so is \( \text{ad}_X \) nilpotent acting on \( \text{End}(V) \).* To see this, suppose \( X \) in \( g \) is nilpotent, say \( X^n = 0 \). Therefore the operators
\[
\Lambda_X: A \mapsto XA \\
R_X: A \mapsto AX
\]
acting on matrices \( A \) satisfy say \( \Lambda_X^n = R_X^n = 0 \).

But these commute with each other. Since \( \text{ad} = \Lambda_X - R_X \), we have therefore
\[
\text{ad}^{2n} = (\Lambda_X - R_X)^{2n} = \sum \binom{2n}{k} (-1)^k L_X^k R_X^{2n-k} = 0,
\]
since either \( k \geq n \) or \( 2n - k \geq n \).

We are given \( g \subseteq \mathfrak{gl}_n \), and want to (a) show \( g^n V = 0 \) for \( n \gg 0 \) or (b) find \( X \) in \( g \) which is not nilpotent.

Choose \( X = X_1 \) in \( g \) (say, the first element of a given basis). Either \( X \) is nilpotent, or it is not. If not, we are through. If it is then (a) the subspace \( U_1 \) of vectors annihilated by \( X \) is not 0, and (b) by the remark just above, \( \text{ad}_X \) is nilpotent. It takes the one-dimensional Lie algebra \( \mathfrak{h} = \mathfrak{h}_1 \subseteq \mathfrak{g} \) spanned by \( X \) into itself, so also acts linearly on the quotient \( \mathfrak{g}/\mathfrak{h} \), and we can find \( X_2 \notin \mathfrak{h} \) in \( g \) such that \( \text{ad}_X(X_2) \) lies in \( \mathfrak{h} \). Thus the space \( \mathfrak{h}_2 \) spanned by \( X_1 \) and \( X_2 \) is a Lie subalgebra of \( g \).

We can continue. At each stage we have a Lie subalgebra \( \mathfrak{h}_i \) of \( g \) together with a non-zero subspace \( U_i \) of \( V \) annihilated by it, and we also have a subalgebra \( \mathfrak{h}_{i+1} \) spanned by \( X_{i+1} \) and \( \mathfrak{h}_i \) with \( [X_{i+1}, \mathfrak{h}_i] \subseteq \mathfrak{h}_i \). The space \( U_i \) is stable under \( \mathfrak{h}_{i+1} \) since for \( X \) in \( \mathfrak{h}_i \)
\[
XX_{i+1}u = X_{i+1}Xu + [X, X_{i+1}]u = 0.
\]

We now set \( U_{i+1} \) equal to the space of all vectors annihilated by \( X_{i+1} \). Either it is non-trivial and we continue on, or \( X_{i+1} \) acting on \( U_i \) is non-nilpotent. There are in the end one of two outcomes: either we find some element of \( g \) that is not nilpotent on \( V \), or \( \mathfrak{h}_n = g \) and we have found a non-zero subspace of \( V \) annihilated by \( g \).

**10. Representations of a nilpotent Lie algebra**

If \( g \) is an abelian Lie algebra and \( F \) is algebraically closed, any finite-dimensional vector space on which \( g \) acts is the sum of certain primary components associated to maximal ideals of the polynomial algebra \( F[\mathfrak{g}] \).

The component associated to the maximal ideal \( \mathfrak{m} \) is the subspace of \( F[\mathfrak{g}] \) annihilated by some power of \( \mathfrak{m} \). This can be proven by a simple induction argument on the dimension of \( \mathfrak{g} \), starting with Fitting's Lemma applied to operators \( T - \lambda I \) for eigenvalues \( \lambda \) of \( T \). This works because if \( (X - \lambda)^m v = 0 \) and \( Y \) commutes with \( X \) then
\[
(X - \lambda)^m Y v = Y (X - \lambda)^m v = 0.
\]

In this section I'll show that something analogous is valid for any nilpotent Lie algebra, although the argument is a bit less simple.

Suppose \( g \) to be a nilpotent Lie algebra and \( V \) to be a \( g \)-module. This means that we are given a map \( \varphi \) from \( g \) to the Lie algebra \( \mathfrak{gl}(V) \), but for the moment I'll assume \( g \) to be embedded into \( \text{End}(V) \). With this assumption, I can ignore \( \varphi \) in notation.

For every \( n \) there exists a linear map
\[
\Phi_n: \bigotimes^n g \otimes V \longrightarrow V, \quad X_1 \otimes \ldots \otimes X_n \otimes v \longmapsto X_1 \ldots X_n v.
\]
Define $g^n V$ to be its image. There also exists a canonical map

$$V \longrightarrow \text{Hom}(\otimes^n g, V)$$

that takes $v$ to the map taking $T$ in $\otimes^n g$ to $\Phi_n(T \otimes v)$. Let $V(g^n)$ be its kernel. The first sequence is weakly decreasing, the second weakly increasing, and they both eventually stabilize. Let $(\langle g \rangle) V$ be the intersection of all the $g^n V$, and $V(\langle g \rangle)$ be the union of the $V(g^n)$.

**10.1. Lemma.** For $n \geq \dim V$

$$V(\langle g \rangle) = \bigcap_{X \in g} V(\langle X \rangle).$$

**Proof.** I’ll show in a moment that each space $V(\langle X \rangle)$ is stable under $g$. Their intersection is then also $g$-stable. It certainly contains $V(\langle g \rangle)$, and is equal to it by Engel’s Criterion.

**10.2. Lemma.** Suppose $X, Y$ to be two endomorphisms of the finite-dimensional vector space $V$. If $\text{ad}^n_X Y = 0$ for some $n$, then the Fitting components of $X$ are stable under $Y$.

**Proof.** Suppose $X^\ell v = 0$. We want to know that for $m \gg 0$ we also have $X^m Y v = 0$. We start out:

$$XY = YX + \text{ad}_X Y$$

$$X^2 Y = X(YX + \text{ad}_X Y)$$

$$= (XY)X + X(ad_X Y)$$

$$= (YX + ad_Y X)X + (ad_Y X)X + ad_Y X$$

$$= YX^2 + 2(ad_Y X)X + ad_Y X,$$

which leads us to try proving by induction that

$$X^m Y = YX^m + m(ad_Y X)X^{m-1} + \cdots + m(ad_Y^{m-1} X)X + ad_Y^m Y$$

$$= \sum_{k=0}^{m} \binom{m}{k} (ad_Y^k X)X^{m-k}.$$

We can make the inductive transition:

$$X^m Y = \sum_{k=0}^{m} \binom{m}{k} (ad_Y^{m-k} X)X^k$$

$$X^{m+1} Y = \sum_{k=0}^{m} \binom{m}{k} (ad_Y^{m-k} X)X^k$$

$$X^{m+1} Y = \sum_{k=0}^{m} \binom{m}{k} ((ad_Y^{m-k} X)X^{k+1} + (ad_Y^{m-k-1} X)X^k)$$

$$= \sum_{k=0}^{m+1} \binom{m+1}{k} (ad_Y^{m-k} X)X^k.$$

Now if $X^\ell v = 0$ and $\text{ad}_X^n Y = 0$ then

$$X^{\ell+n} Y v = \sum_{k=0}^{\ell+n} \binom{\ell+n}{k} (ad_Y^k X)X^{\ell+n-k} v = 0.$$
Introduction to Lie algebras 21

since in the sum either \( k \geq n \) or \( \ell + n - k > \ell \). A similar argument will show that \( V^{k} \) is \( \mathfrak{g} \)-stable, once we prove

\[
Y^{m}X^{m} = X^{m}Y - mX^{m-1}(ad_{X}Y) + \cdots \pm mX(ad_{X}^{m-1}Y) \mp ad_{X}^{m}Y.
\]

Here is a nilpotent extension of Fitting’s Lemma:

10.3. Proposition. Any \( \mathfrak{g} \)-module \( V \) is the direct sum of \( V^{(\mathfrak{g})} \) and \( (\mathfrak{g})V \).

Proof. We proceed by induction on \( \dim V \). If every \( X \) in \( \mathfrak{g} \) acts nilpotently then by Engel’s Criterion \( V = V^{(\mathfrak{g})} \) and \( (\mathfrak{g})V = 0 \). Otherwise, there exists \( X \) in \( \mathfrak{g} \) with \( (\mathfrak{g})V \neq 0 \). If \( (\mathfrak{g})V = V \) then necessarily \( V^{(\mathfrak{g})} = 0 \) and we are through. Otherwise we have the decomposition

\[
V = V^{(\mathfrak{g})}V \oplus (\mathfrak{g})V
\]

into two proper subspaces. By Lemma 10.2 each of these is stable under \( \mathfrak{g} \), and the second is contained in \( (\mathfrak{g})V \). Let the first be \( U \). We may apply the induction hypothesis to decompose it into a sum \( U^{(\mathfrak{g})} \oplus (\mathfrak{g})U \). This gives us in turn the decomposition

\[
V = U^{(\mathfrak{g})} \oplus (\mathfrak{g})U \oplus (\mathfrak{g})V
\]

But the first is the same as \( V^{(\mathfrak{g})} \) and the second is contained in \( (\mathfrak{g})V \). On the other hand, if \( U = V^{(\mathfrak{g})} \) and \( \mathfrak{g}^{n}U = 0 \) then

\[
\mathfrak{g}^{n}V \subseteq (\mathfrak{g})U \oplus (\mathfrak{g})V
\]

so in fact

\[
(\mathfrak{g})V = (\mathfrak{g})U \oplus (\mathfrak{g})V.
\]

This might prove useful sometime:

10.4. Corollary. The functors

\[
V \Rightarrow V^{(\mathfrak{g})}
\]

\[
V \Rightarrow (\mathfrak{g})V
\]

are both exact.

Proof. From the Proposition, since the first is clearly left exact and the second right exact.

I must now take into account the Lie homomorphism \( \varphi \) from \( \mathfrak{g} \) to \( \text{End}(V) \). If \( \lambda \) is a character of \( \mathfrak{g} \), we can use it to define a ‘twisted’ representation of \( \mathfrak{g} \) on \( V \) according to the formula

\[
\varphi - \lambda I : X \mapsto \varphi(X) - \lambda(X)I.
\]

This is a representation because

\[
[\varphi(X) - \lambda(X)I, \varphi(Y) - \lambda(Y)I] = [X, Y] = [X, Y] - \lambda([X, Y])
\]

since \( \lambda([X, Y]) = 0 \) by definition of character. In this new representation, the vectors annihilated by \( \mathfrak{g}^{n} \) are those annihilated by all products

\[
\prod_{i=1}^{n} (\varphi(X_{i}) - \lambda(X_{i})I)
\]

in the original.

Assume for the rest of this section that \( F \) is algebraically closed.

If we apply the previous Proposition successively to the \( \mathfrak{g} \)-module \( \mathfrak{g}^{n}V \), we get a decomposition of \( V \) into a direct sum of components \( V^{(\mathfrak{g}, \lambda)} \) annihilated by such products, given this:
10.5. Lemma. If $V$ is a finite-dimensional module over the nilpotent Lie algebra $\mathfrak{g}$, there exists at least one eigenvector for it.

Proof. I am going to prove something a bit more general that will be useful later on.

10.6. Lemma. Suppose $\mathfrak{g}$ to be a Lie algebra possessing an increasing sequence of Lie subalgebras

$$\mathfrak{g}_0 = 0 \subset \mathfrak{g}_1 \subset \ldots \subset \mathfrak{g}_n = \mathfrak{g}$$

with the property that $D_{\mathfrak{g}_{n+1}} \subseteq D_{\mathfrak{g}_n}$. Any finite-dimensional module over $\mathfrak{g}$ possesses an eigenvector for it.

The proof of this is by induction on the dimension $m$ of $\mathfrak{g}$. The case that $m = 1$ is trivial, so suppose $m > 1$. The Lemma will follow by induction from this new Lemma:

10.7. Lemma. Suppose $\mathfrak{h} \subset \mathfrak{g}$ of codimension one, with $\mathfrak{h} \subseteq D_{\mathfrak{g}}$. Any eigenspace for $\mathfrak{h}$ is stable under $\mathfrak{g}$.

An eigenvector of all of $\mathfrak{g}$ is one for $\mathfrak{h}$, so this is a very optimistic hope about a possible converse.

Proof. Let $X \not= 0$ be an element of $\mathfrak{g}$ not in $\mathfrak{h}$. Let $v \not= 0$ be an eigenvector of $\mathfrak{h}$, so that

$$H \cdot v = \lambda(H)v$$

for all $H$ in $\mathfrak{h}$, where $\lambda$ is a character of $\mathfrak{h}$. We would like to know that all $v_k = X^kv$ are also eigenvectors for $\mathfrak{h}$. At any rate, let $V_m$ be the space spanned by the vectors $v_k$ for $k \leq m$. Since $XV_m \subseteq V_{m+1}$, the union of these is at least a finite-dimensional space $V$, taken into itself by $X$. I shall show that each space $V_m$ is stable under $\mathfrak{h}$, and that the representation of $\mathfrak{h}$ on $V_m/V_{m-1}$ is by $\lambda$. I do this by induction. It is true for $m = 0$ by assumption.

For $H$ in $\mathfrak{h}$ and $m \geq 1$ we have

$$(H - \lambda(H))v_m = (H - \lambda(H))Xv_{m-1}$$

$$= HXv_{m-1} - \lambda(H)Xv_{m-1}$$

$$= XHv_{m-1} + [H, X]v_{m-1} - X\lambda(H)v_{m-1}$$

$$= X(H - \lambda(H))v_{m-1} + [H, X]v_{m-1}.$$  

But by induction $(H - \lambda(H))v_{m-1}$ lies in $V_{m-2}$, so the first term lies in $V_{m-1}$. But, again by induction, since $[H, X]$ lies in $\mathfrak{h}$ the third term also lies in $V_{m-1}$.

We now therefore know that $V_m$ is stable with respect to both $X$ and $\mathfrak{h}$, hence all of $\mathfrak{g}$, and we also know that $H - \lambda(H)$ is nilpotent on it for every $H$ in $\mathfrak{h}$.

I now claim that in fact every vector in $V_m$ is an eigenvector for $\mathfrak{h}$. First of all, the trace of any $H$ in $\mathfrak{h}$ is equal to $d_\mathfrak{g} \cdot \lambda(H)$, where $d_\mathfrak{g}$ is the dimension of $V_m$. On the other hand, the trace of $[H, X] = HX - XH$ is 0. Therefore $\lambda([H, X]) = 0$ for every $H$ in $\mathfrak{h}$. But if we assume that $Hv_{m-1} = \lambda v_{m-1}$ for all $H$ in $\mathfrak{g}$ then

$$Hv_m = HXv_{m-1} = XHv_{m-1} + [H, X]v_{m-1} = \lambda v_m + \lambda([H, X])v_{m-1} = \lambda v_m.$$  

Any eigenvector for $X$ in $V_m$ will be an eigenvector for all of $\mathfrak{g}$. 

In summary:

10.8. Theorem. If $V$ is a module over the nilpotent Lie algebra $\mathfrak{g}$, there exists a finite set $\Lambda$ of characters $\lambda$ of $\mathfrak{g}$ and a primary decomposition

$$V = \bigoplus_{\mathfrak{g}, \lambda} V(\mathfrak{g}, \lambda).$$

These are called the primary components of the $\mathfrak{g}$-module $V$.

We’ll see later that $\mathfrak{g}$ is embedded into its universal enveloping algebra $U(\mathfrak{g})$, an associative algebra which it generates. Every $\mathfrak{g}$-module is automatically a module over $U(\mathfrak{g})$. If $\mathfrak{g}$ is abelian, its universal enveloping algebra is a polynomial algebra. This Proposition can be seen as an indication that basic results about modules over polynomial algebras extend to the universal enveloping algebras of nilpotent Lie algebras. For more along these lines see [McConnell:1967] and [Gabriel-Nouazé:1967].
11. Cartan subalgebras

If \( g \) is a Lie algebra, the **normalizer** of any subalgebra \( h \) of \( g \) is the space of all \( X \) such that \( \text{ad}Xh \subseteq h \). It is a Lie algebra that contains \( h \) itself as an ideal.

A **Cartan subalgebra** is a nilpotent subalgebra whose normalizer is itself. It is not completely obvious that Cartan subalgebras exist, but in fact:

11.1. **Theorem.** Every Lie algebra \( g \) possesses at least one Cartan subalgebra.

**Proof.** As [Merrin:1994] points out, a constructive proof is an easy consequence of Engel’s Criterion. Recall that for every \( X \) in \( g \)

\[
\mathfrak{g}(\text{ad}X) = \{ Y \in \mathfrak{g} \mid \text{ad}^n_XY = 0 \text{ for some } n > 0 \}. 
\]

11.2. **Lemma.** (Leibniz’ rule) If \( D \) is a derivation of a Lie algebra then

\[
D^n[A, B] = \sum_{k=0}^{n} \binom{n}{k} [D^kA, D^{n-k}B].
\]

**Proof.** By a straightforward induction.

As a consequence of Leibniz’ rule, \( \mathfrak{g}(\text{ad}X) \) is a Lie subalgebra of \( \mathfrak{g} \).

Suppose \( Y \) to lie in the normalizer of \( \mathfrak{g}(\text{ad}X) \). Since \( X \) itself lies in \( \mathfrak{g}(\text{ad}X) \) we have

\[
[X, Y] = -[Y, X] \in \mathfrak{g}(\text{ad}X),
\]

and hence \( Y \) itself will lie in \( \mathfrak{g}(\text{ad}X) \). So the normalizer of \( \mathfrak{g}(\text{ad}X) \) is itself. It will be a Cartan subalgebra if and only if it is nilpotent. In this case, for the moment, \( X \) will be called regular.

Therefore in order to prove Theorem 11.1 it suffices to prove that there exists a regular \( X \) in \( \mathfrak{g} \).

We start off with an arbitrary element \( X \) in \( \mathfrak{g} \). I shall show that if \( \mathfrak{g}(\text{ad}X) \) is not nilpotent, there exists \( Z \) in \( \mathfrak{g}(\text{ad}X) \) such that \( \mathfrak{g}(\text{ad}Z) \) is a proper subspace of \( \mathfrak{g}(\text{ad}X) \), or in other words that \( \alpha A_X + \beta A_Y \) is not nilpotent but \( \alpha C_X + \beta C_Y \) is still invertible. Let \( \alpha, \beta \) be variables. Consider the characteristic polynomial

\[
D_M(t, \alpha, \beta) = \det(tI - \alpha A_X - \beta A_Y) = \det(tI - \alpha A_X - \beta A_Y) = D_A(t, \alpha, \beta)D_C(t, \alpha, \beta). 
\]

If \( k = \text{dim} \mathfrak{g}(\text{ad}X) \), then \( t^k \) cannot divide \( D_A(t, \alpha, \beta) \) because this would contradict the fact that \( A_Y \) is not nilpotent. And \( t \) cannot divide \( D_C(t, \alpha, \beta) \) since \( C_X \) is invertible. Thus if

\[
D_M(t, \alpha, \beta) = \sum t^k c_k(\alpha, \beta)
\]
then some $c_\ell$ with $\ell < k$ must not vanish identically as a polynomial in $\alpha, \beta$. Choose numbers $\alpha, \beta$ so it doesn’t vanish. In these circumstances, $g(\text{ad}_Z)$ is a proper subspace of $g(\text{ad}_X)$. Induction allows us to conclude.

If $\mathfrak{h}$ is a Cartan subalgebra of $g$ then according to Theorem 10.8 $g$ decomposes into a direct sum of components $g(\text{ad}_{h, \lambda})$. The characters $\lambda$ that occur are called the roots of $g$ with respect to $h$. The algebra itself is certainly contained in the root space $g(\text{ad}_h)$.

11.3. Proposition. If $h$ is a nilpotent Lie subalgebra of $g$, the following are equivalent:
(a) it is a Cartan subalgebra;
(b) it is the same as $g(\text{ad}_h)$.

Proof. Suppose $h$ nilpotent, let $n$ be its normalizer in $g$, and let $\tau = g(\text{ad}_h)$. If $Y \in n$ then for any $X \in h$

$$[X, Y] = -[Y, X] \in h$$

and $\text{ad}_X^n Y = 0$ for some $n$, so

$$h \subseteq n \subseteq \tau.$$ 

Thus if $h = \tau$ then $h = n$ and $h$ is a Cartan subalgebra.

Conversely, suppose $h = n$. Then $h$ acts nilpotently on $\tau = \tau/h$, so by Engel’s Theorem either $\tau = \{0\}$ or there exists $Y \neq 0$ such that $\text{ad}_X Y = 0$ for all $X$ in $h$. But if $Y$ in $\tau$ has image $Y'$ then $Y'$ must lie in $n$, a contradiction.

11.4. Proposition. For characters $\lambda, \mu$ of a Cartan subalgebra

$$[g_\lambda, g_\mu] \subseteq g_{\lambda + \mu}.$$ 

Proof. Suppose $X$ in $a, U$ in $g_\lambda$, $Y$ in $g_\mu$. Let $D = \text{ad}_X$. The case most commonly seen is that in which the Cartan subalgebra is abelian and the root spaces are eigenspaces. Since $D$ is a derivation, we have

$$(D - \lambda - \mu)[X, Y] = [(D - \lambda)X, Y] + [X, (D - \mu)Y] = 0,$$

which proves the proposition in this simplest case. But the formula

$$(D - \lambda - \mu)[X, Y] = [(D - \lambda)X, Y] + [X, (D - \mu)Y]$$

is valid for any $X, Y$ at all. We may apply it repeatedly to get

$$(D - \lambda - \mu)[X, Y] = \sum_{p=0}^{n} \binom{n}{p} [(D - \lambda)^p X, (D - \mu)^{n-p} Y],$$

from which the Proposition follows if we just take $n$ large enough.
12. Conjugacy of Cartan subalgebras

In this section, let \( \mathfrak{g} \) be a Lie algebra defined over the algebraically closed field \( F \) (still of characteristic 0). Recall that \( \mathfrak{g} \) decomposes into root spaces with respect to any Cartan subalgebra.

If \( x \) is a nilpotent element of \( \mathfrak{g} \), and in particular if it is in some non-trivial root space, the exponential \( \exp(\text{ad}_x) \) is an automorphism of \( \mathfrak{g} \). Let \( \text{Int}(\mathfrak{g}) \) be the group generated by these. In this section I’ll prove, along with other related results:

12.1. Theorem. Any two Cartan subalgebras of \( \mathfrak{g} \) are conjugate by an element of \( \text{Int}(\mathfrak{g}) \).

I follow the treatment attributed in [Serre:1966] to Chevalley and presented in [Cartier:1955].

The proof starts with a simple, useful (and perhaps well known) result from algebraic geometry. It is a partial substitute for the implicit function theorem, which does not remain valid.

12.2. Lemma. Suppose \( \varphi: U \to V \) to be a polynomial map from one vector space over \( F \) to another. If \( d\varphi \) is surjective at some point of \( U \), then \( \varphi \) is generically surjective.

The precise conclusion is that

\[
\text{if we are given a polynomial } P \text{ on } U, \text{ we can find a polynomial } Q \text{ on } V \text{ with the property that every point } v \text{ of } V \text{ with } Q(v) \neq 0 \text{ is equal to } \varphi(u) \text{ for some } u \text{ in } U \text{ with } P(u) \neq 0.
\]

Intuitively, this is quite plausible. Algebraic closure is certainly necessary, as the map \( x \mapsto x^2 \) from \( \mathbb{R} \) to itself shows.

Proof. Say \( U \) has dimension \( m \), \( V \) has dimension \( n \). Let \( (x_i) \) be the coordinate system on \( U \), \( (y_j) \) that on \( V \). Under the hypothesis of the Lemma, and by a change of coordinates if necessary, we may assume that \( \varphi \) is given by the formula

\[
y_j = x_j + \text{second order terms for } 1 \leq j \leq m.
\]

Step 1. I first show that composition with \( f \), which maps \( F[V] \) to \( F[U] \), is an injection. Intuitively, this means that the image of \( U \) in \( V \) is not a proper subvariety of \( V \), since if \( I \) is the ideal of \( F[V] \) vanishing on the image, the map from \( F[V] \) to \( F[U] \) is filtered through \( F[V]/I \). Given \( f(y) \neq 0 \), suppose it has order \( N \) at 0, so

\[
f = \sum_{|k|=N} f_k y^k + \text{terms of order } > N
\]

with some \( f_k \neq 0 \). But then

\[
f(\varphi(x)) = \sum_{|k|=N} f(k) x^k + \text{terms of order } > N.
\]

Hence the image of \( f \) in \( F[U] \) does not vanish.

Step 2. From this point on, we do not need the assumption about \( d\varphi \) any more—we just need the ring \( F[V] \) to be embedded in \( F[U] \). A point of \( U \) may be identified with a ring homomorphism \( \pi_u \) from \( F[U] \) to \( F \), and similarly for a point of \( V \). If \( P \) is a polynomial in \( F[U] \) and a \( u \) point in \( U \) then \( P(u) = 0 \) if and only if \( \pi_u(P) = 0 \). Thus what must be proven is this:

Given \( P \) in \( F[U] \), we can find \( Q \) in \( F[V] \) such that any homomorphism \( \pi: F[V] \to F \) with \( \pi(Q) \neq 0 \) may be extended to a ring homomorphism \( \Pi: F[U] \to F \) with \( \Pi(P) \neq 0 \).

Define by recursion

\[R_0 = F[V], \quad R_k = R_{k-1}[x_k].\]

Thus \( R_m = F[U] \). An easy induction reduces the proof of the claim to this:
12.3. Lemma. Suppose $R$ to be an integral domain containing $F$, $S = R[s]$ an integral domain containing $R$ and generated over $R$ by a single element $s$. If $\sigma \neq 0$ is an element of $S$, there exists an element $\rho$ in $R$ with the following property: every homomorphism $\pi: R \to F$ with $\pi(R) \neq 0$ lifts to a homomorphism $\pi_S: S \to F$ with $\pi_S(\sigma) \neq 0$.

Proof of the Lemma. Represent $S$ as $R[x]/I$, where $x$ is a variable and $I$ an ideal. Because $S$ contains $R$, we must have $R \cap I = (0)$. There are two cases:

(a) $I = (0)$ and $S = R[x]$. Suppose $\sigma = \sum r_i x^i$. If $\pi: r \mapsto \pi$ is a ring homomorphism, the extensions to $S$ correspond to a choice of image $\pi$ of $x$, which is arbitrary. The image of $\sigma$ is $\sum \pi_i x^i$. Since $\sigma \neq 0$ we must have some $r_k \neq 0$ in $R$. Set $\rho = r_k$. If $\pi \neq 0$ the polynomial $\sum \pi_i x^i$ does not vanish identically, and only has a finite number of roots. Since $F$ is algebraically closed it is infinite, and we may choose $\pi$ in $F$ such that $\sum \pi_i x^i \neq 0$.

(b) $I \neq (0)$. If $\sum q_i x^i$ lies in $I$, then some $q_i \neq 0$ with $i > 0$ since $I \cap R = (0)$. Let $P(x) = \sum p_i x^i$ be of minimal degree in $I$. If $Q(x)$ is any other polynomial in $I$ then the division algorithm gives

$$p_n^0Q(x) = U(x)P(x) + V(x)$$

with the degree of $V$ less than $d$. But then $V(x)$ must also lie in $I$, hence $V(x) = 0$. Because $S$ is an integral domain, $I$ is a prime ideal, and since $I \cap R = (0)$, we conclude:

A polynomial $Q(x)$ lies in $I$ if and only if $p_n^0Q(x)$ is a multiple of $P(x)$ in $R[x]$ for some $n$.

Suppose $r \mapsto \pi$ is a ring homomorphisms from $R$ to $F$, and let $m$ be its kernel, a maximal ideal of $R$. Extensions to $S$ amount to homomorphisms from $S = F/mS$ to $F$, or in other words to maximal ideals of $S$ to $F$. If $\overline{\pi}_d \neq 0$, the quotient $S$ is the same as $F[x]/(\overline{\pi})$. Suppose $P$ is the image in $S$ of the polynomial $\Pi = \sum r_i x^i$. Apply the division algorithm to it to get

$$p_n^0\Pi(x) = U(x)P(x) + V(x)$$

with the degree of $V$ less than $d$. Because $P \neq 0$ in $S$, $\Pi(x)$ is not in $I$ and $V(x) \neq 0$. Let $v$ be one of its coefficients in $R$. Now set $Q = p_n^0 v$. If $\overline{Q} \neq 0$ then the image $\overline{P}$ of $\Pi$ in $S$ is not 0. Since $F$ is algebraically closed, we may find a homomorphism from $\overline{S}$ to $F$ that does not annihilate $\overline{P}$.

I now turn to the original question of conjugacy of Cartan subalgebras in a Lie algebra. This turns out to be closely related to other important properties of Cartan subalgebras that I shall prove at the same time.

If $A$ is any linear transformation of a finite-dimensional vector space of dimension $n$ over an algebraically closed field, its characteristic polynomial $\det(T - A)$ will be of the form $\delta_1 T^\ell + \cdots + T^n$ where $\ell$ is least such that $\delta_1 \neq 0$. The integer $\ell$ is called the nilpotent rank of $A$. The rank of a Lie algebra is the smallest nilpotent rank of all linear transformations $ad_X$ for $X$ in $g$. It is at most $n$ since the coefficient of $T^n$ is 1, and it is at least 1 since $[X, X] = 0$. An element $X$ is called regular if its nilpotent rank is the same as that of $g$. If $g$ is assigned a coordinate system the coefficients of the characteristic polynomial of $ad_X$ will be a polynomial in the coefficients of $X$, so the regular elements are the complement in $g$ of a proper algebraic subvariety.

It will follow from Theorem 12.1 that all Cartan algebras have the same dimension. More precisely:

12.4. Proposition. All Cartan subalgebras have dimension equal to the rank of $g$.

12.5. Proposition. Suppose $a$ to be a Cartan subalgebra of $g$, and suppose $x$ to lie in $a$. Then $(\lambda, x) \neq 0$ for all roots $\lambda$ if and only if $x$ is regular in $g$.

I’ll prove these and Theorem 12.1 all at the same time. Let $\ell$ be the rank of $g$, $\delta_\ell$ the corresponding coefficient of the characteristic polynomials, so that $\delta_\ell(X) \neq 0$ if and only if $X$ is regular. Fix for the moment the Cartan
subalgebra \( a \) and the associated root decomposition of \( g \). Let \( g_{\neq 0} \) be the direct sum of the \( g_\lambda \) with \( \lambda \neq 0 \). Choose a basis \( (X_i) \) (say for \( 1 \leq i \leq n \)) of \( g_{\neq 0} \) with \( X_i \) in \( g_{\lambda_i} \), and define the map

\[
\varphi: \left( \bigoplus_{\lambda \neq 0} g_\lambda \right) \oplus a \longrightarrow g
\]

taking

\[
(X_i) \times H \longmapsto \exp(\text{ad}_{X_i}) \ldots \exp(\text{ad}_{X_n}) \cdot H.
\]

Since

\[
\exp(\text{ad}_{tX}) Y = I + \text{ad}_X Y + \cdots = I + t[X, Y] + \cdots
\]

the Jacobian map \( d\varphi \) at \( h \) in \( a \) takes \( (X_i) \oplus H \) to \( (\langle \lambda_i, h \rangle) X_i \oplus H \). It is thus an isomorphism of \( U = V = g \) with itself as long as \( \langle \lambda_i, h \rangle \neq 0 \) for all \( i \). Let \( P \) be the product \( \prod \lambda(H) \) on \( ((t, X_i), H) \), and let \( a' \) be the set of \( H \) in \( a \) with \( P(H) \neq 0 \). By the Proposition, there exists a polynomial \( Q(x) \) on \( g \) with the property that if \( Q(X) \neq 0 \) then \( X \) is conjugate by \( \text{Int}(g) \) to some \( H \) in \( a' \). The set of regular elements of \( g \) is Zariski-open in \( g \) and invariant under \( \text{Int}(g) \), so any \( X \) in \( g \) with \( Q(X)P(X) \neq 0 \) is \( f(g) \)-conjugate to some regular \( H \) in \( a \). But for such an \( H \), the space on which \( \text{ad}_H \) is nilpotent is all of \( a \) on the one hand, and of dimension \( \ell \) on the other. Hence \( a \) has dimension \( \ell \).

If \( b \) is another Cartan subalgebra, the same holds for it, and there must exist some \( x \) in \( g \) which is at once conjugate to a regular element of \( a \) as well as one of \( b \), which implies that the nilpotent component of \( x \) is conjugate to \( a \) as well as \( b \), so they must also be conjugate to each other.

13. Killing forms

If \( \rho \) is a (finite-dimensional) representation of \( g \), we can define a symmetric inner product

\[
K_\rho: g \otimes g \rightarrow F; \quad X \otimes Y \mapsto \text{trace}(\rho(X)\rho(Y)).
\]

It is called the Killing form associated to \( \rho \). If \( \rho \) is not specified, it is assumed to be the adjoint representation \( \text{ad} \).

13.1. Proposition. For any representation \( \rho \) the bilinear form \( K_\rho \) is \( g \)-invariant. If \( g \) is the Lie algebra of \( G \) and \( \rho \) is the differential of a smooth representation of \( G \) then it is also \( G \)-invariant.

The first assertion means that the adjoint representation of \( g \) on itself takes \( g \) into the Lie algebra of the orthogonal group preserving this bilinear form.

Proof. First I’ll show invariance under \( g \). We must show that

\[
K_\rho(\text{ad}_X Y, Z) + K_\rho(Y, \text{ad}_X Z) = 0.
\]

This translates to the condition

\[
\text{trace}(\rho(\text{ad}_X Y)\rho(Z)) + \text{trace}(\rho(Y)\rho(\text{ad}_X Z)) = 0
\]

and then to

\[
\text{trace}(\rho([X, Y]\rho(Z)) + \text{trace}(\rho(Y)\rho([X, Z])) = \text{trace}(\rho(X)\rho(Y)\rho(Z) - \rho(Y)\rho(X)\rho(Z) + \rho(Y)\rho(X)\rho(Z) - \rho(Y)\rho(Z)\rho(X))
\]

\[
= \text{trace}(\rho(X)\rho(Y)\rho(Z) - \rho(Y)\rho(Z)\rho(X)) = 0.
\]
Now the matter of $G$-invariance. For ease of reading I'll write $\text{Ad}(g)X$ as $gXg^{-1}$. Then
\[
\text{trace}(\rho(gXg^{-1})\rho(gYg^{-1})) = \text{trace}(\rho(g)\rho(X)\rho(g^{-1})\rho(g)\rho(Y)\rho(g^{-1}))
\]
\[
= \text{trace}(\rho(g)\rho(X)\rho(Y))
\]
\[
= \text{trace}(\rho(X)\rho(Y)) .
\]

The Killing form on $\mathfrak{g}$ itself is characteristic.

13.2. Proposition. The Killing form of $\mathfrak{g}$ is invariant under any automorphism of $\mathfrak{g}$.

Proof. If $\alpha$ is an automorphism of $\mathfrak{g}$ then by definition $[\alpha(X), \alpha(Y)] = \alpha([X, Y])$. In other words
\[
ad_{\alpha(X)} = \alpha \ad_X \alpha^{-1}.
\]
But then
\[
\text{trace}(\ad_{\alpha(X)}\ad_{\alpha(Y)}) = \text{trace}(\alpha \ad_X \alpha^{-1} \cdot \alpha \ad_Y \alpha^{-1})
\]
\[
= \text{trace}(\alpha \ad_X \ad_Y \alpha^{-1})
\]
\[
= \text{trace}(\ad_X \ad_Y) .
\]

If $K$ is a $\mathfrak{g}$-invariant bilinear form on $\mathfrak{g}$, its radical is the subspace of $X$ in $\mathfrak{g}$ such that $K(X, g) = 0$.

13.3. Proposition. The radical of any $\mathfrak{g}$-invariant bilinear form is an ideal of $\mathfrak{g}$.

Proof. Immediate.

13.4. Proposition. If $\mathfrak{g} = \mathfrak{sl}_n$ then $K_{\text{ad}}(X, Y) = 2n \text{trace}(XY)$.

In particular, it is non-degenerate.

Proof. The Lie algebra $\mathfrak{sl}_n$ is the space of $n \times n$ matrices of trace 0, with
\[
ad_X(Y) = XY - YX .
\]
This action extends to the space of all matrices, and the action on the complement of $\mathfrak{sl}_n$ is trivial. Therefore the Killing form $K_{\text{ad}}$ is also that associated to the adjoint action on $\mathfrak{gl}_n = M_n$. So for any $X$ and $Y$ in $M_n$ the Killing form is $\text{trace}(\ad_X \ad_Y)$, where $\ad_X \ad_Y$ takes
\[
Z \mapsto YZ - ZY \mapsto X(YZ - ZY) - (YZ - ZY)X = XYZ - XZY - YZX + ZYX .
\]
If $A$ is any matrix, the trace of each of the maps
\[
Z \mapsto AZ, \quad Z \mapsto ZA
\]
is $n \text{trace}(A)$, since as a left or right module over $M_n$ the space $M_n$ is a sum of $n$ copies of the module $F$. Therefore the trace of the first and last maps is $2n \text{trace}(XY)$. The trace of each of the middle two $Z \mapsto XZY$, $Z \mapsto YZX$ is the product $\text{trace}(X)\text{trace}(Y)$, hence 0 on $\mathfrak{sl}_n$. 


14. Solvable Lie algebras

The derived idea $D\mathfrak{g}$ of $\mathfrak{g}$ is that spanned by all commutators $[X, Y]$ (see Proposition 8.1). It is an ideal in $\mathfrak{g}$, and the quotient $\mathfrak{g}/D\mathfrak{g}$ is abelian. The derived series is that of the $D^n\mathfrak{g}$ where

$$D^0 = \mathfrak{g}, \quad D^{i+1} = [D^i, D^i].$$

Each $D^n$ is an ideal in $D^{n-1}$, but not necessarily an ideal of $\mathfrak{g}$. A Lie algebra is said to be **solvable** if some $D^n = 0$. The ultimate origin of this term is presumably Galois’ criterion for solvability of equations by radicals in terms of what we now call solvable Galois groups.

**14.1. Lemma.** Suppose $E$ a field extension of $F$. The Lie algebra $\mathfrak{g}$ over $F$ is solvable if and only $\mathfrak{g} \otimes_F E$ is.

**Proof.** Since

$$D\mathfrak{g} \otimes_F E = D(\mathfrak{g} \otimes_F E).$$

A Lie algebra is solvable if and only if it possesses a filtration by subalgebras $\mathfrak{g}_i$ with each $\mathfrak{g}_i$ an ideal in $\mathfrak{g}_{i-1}$ and $\mathfrak{g}_{i-1}/\mathfrak{g}_i$ abelian. Any quotient or subalgebra of a solvable Lie algebra is solvable. Conversely, if a Lie algebra is an extension of solvable Lie algebras, it is solvable (this is not true for nilpotent Lie algebras). Every nilpotent Lie algebra is solvable, but not conversely. The prototypical solvable algebra is the Lie algebra $\mathfrak{b}_n$ of all upper triangular $n \times n$ matrices. In this case $D\mathfrak{b}_n$ is $\mathfrak{n}_n$.

**14.2. Theorem.** (Lie’s Criterion) Assume $F$ algebraically closed. A Lie subalgebra $\mathfrak{g}$ of $\mathfrak{gl}_n$ is solvable if and only if some conjugate lies in $\mathfrak{b}_n$.

Algebraic closure of $F$ is necessary. For example, if $F = \mathbb{Q}$ and $\mathfrak{g}$ is the Lie algebra of the multiplicative group of an algebraic extension of $F$ of degree $n$, it is commutative, hence solvable, but not diagonalizable in $\mathfrak{gl}_n(F)$.

**Proof.** An induction argument reduces the proof to showing that there exists a common eigenvector for all $X$ in $\mathfrak{g}$. But this we have seen already in Lemma 10.6.

**14.3. Corollary.** A Lie algebra is solvable if and only if $D\mathfrak{g}$ is nilpotent.

**Proof.** One way is trivial. For the other, apply the Proposition to the image of $\mathfrak{g}$ in $\mathfrak{gl}(\mathfrak{g})$, after extending the base field to an algebraic closure of $F$.

**14.4. Theorem.** (Cartan’s Criterion) If the Lie algebra $\mathfrak{g}$ is solvable, then $\text{trace}(\text{ad}_X \text{ad}_Y) = 0$ for all $X$ in $\mathfrak{g}, Y$ in $D\mathfrak{g}$. Conversely, if $\text{trace}(\text{ad}_X \text{ad}_Y) = 0$ for all $X, Y$ in $D\mathfrak{g}$ then $\mathfrak{g}$ is solvable.

I recall that the inner product $\text{trace}(\text{ad}_X \text{ad}_Y)$ is the Killing form of $\mathfrak{g}$.

**Proof.** Because of Theorem 14.2 the first claim is easy, since if $b$ is an upper triangular matrix and $n$ a nilpotent upper triangular matrix then $nb$ is nilpotent.

As for the second, I follow the argument of [Jacobson:1962]. It was primarily for this purpose that I also followed him in speaking of the primary decomposition of modules over nilpotent Lie algebras.

It suffices to show that $D\mathfrak{g}$ is nilpotent or, by Engel’s Criterion, that each $X$ in $D\mathfrak{g}$ is nilpotent. This is clearly true if $X$ lies in $\mathfrak{g}_\lambda$ with $\lambda \neq 0$, since then $\text{ad}_X \mathfrak{g}_\mu \subseteq \mathfrak{g}_{\lambda + \mu}$ and some $\mathfrak{g}_{\lambda + n\mu} = 0$, so we may assume $X$ to lie in $\mathfrak{g}_0 \cap D\mathfrak{g}$, hence that $X = [X_\lambda, X_{-\lambda}]$ with $X_{\pm \lambda} \in \mathfrak{g}_{\pm \lambda}$.

**14.5. Lemma.** If $X = [X_\lambda, X_{-\lambda}]$ then for every weight $\mu$ of $X$ on any $\mathfrak{g}$-module $V$, $\mu(X)$ is a rational multiple of $\lambda(X)$.

**Proof.** Given $\mu$, the sum of weight spaces $\mathfrak{g}_{\mu + n\lambda}$ is stable under the operators $X_{\pm \lambda}$, and therefore the trace of $X = X_\lambda X_{-\lambda} - X_{-\lambda} X_\lambda$ is $0$ on this space. But if $n_\mu$ is the dimension of $\mathfrak{g}_\mu$, we therefore have

$$\sum_k n_{\mu + k\lambda}(\mu(X) + k\lambda(X)) = 0, \quad \left(\sum_k n_{\mu + k\lambda}\right)\mu(X) = -\lambda(X)\left(\sum_k kn_{\mu + k\lambda}\right).$$
To finish the proof of Cartan’s criterion, say \( \mu(X) = c_{\mu, \lambda}(X) \) for all weights \( \mu \). Then

\[
0 = \text{trace} \ X^2 = \sum_{\mu} n_{\mu} \mu^2(X) = \lambda^2(X) \left( \sum n_{\mu} c^2_{\mu, \lambda} \right)
\]

from which we conclude \( \lambda(X) = 0 \), and this then implies that \( \text{ad}_X \) is nilpotent.

15. The universal enveloping algebra

Roughly speaking, the universal enveloping algebra \( U(\mathfrak{g}) \) of a Lie algebra \( \mathfrak{g} \) is the associative algebra generated by the space \( \mathfrak{g} \) subject to the relations \( XY - YX = [X, Y] \) for \( X, Y \) in \( \mathfrak{g} \). More formally, it is the quotient of the tensor algebra \( \bigotimes \mathfrak{g} \) modulo these relations, in other words by the ideal \( I = I_\mathfrak{g} \) of the tensor algebra spanned by all

\[
A \otimes Y \otimes Z \otimes B - A \otimes Z \otimes Y \otimes B - A \otimes [Y, Z] \otimes B.
\]

The following is immediate from the definition:

**15.1. Theorem.** The universal enveloping algebra is universal in the sense that any linear map \( \varphi \) from \( \mathfrak{g} \) to an associative algebra \( A \) such that

\[
\varphi([X, Y]) = \varphi(X) \varphi(Y) - \varphi(Y) \varphi(X)
\]

determines a unique ring homomorphism from \( U(\mathfrak{g}) \) extending \( \varphi \).

Choose a basis \( \Xi = \{X_i\} \) of \( \mathfrak{g} \), ordered by index. The monomials

\[
X_{i_1} \otimes \ldots \otimes X_{i_n}
\]

form a basis of \( \bigotimes \mathfrak{g} \). I call a monomial **reduced** if \( i_k \leq i_\ell \) whenever \( k \leq \ell \).

**15.2. Lemma.** (Poincaré-Birkhoff-Witt) The image in \( U(\mathfrak{g}) \) of the ordered monomials in \( \bigotimes \mathfrak{g} \) are a basis.

**Proof.** It will be long. I’ll call a tensor **reduced** if all of the monomials occurring in it with non-zero coefficients are reduced. We want to show that for every tensor \( x \) there exists a unique tensor \( \text{red}(x) \) such that \( x - \text{red}(x) \) lies in \( I \).

It is relatively easy to see that ordered monomials span the enveloping algebra. I need for this a measure of how non-reduced a tensor is. Define the degree of a monomial in \( \bigotimes \mathfrak{g} \):

\[
\deg(Y_1 \otimes \ldots \otimes Y_n) = n.
\]

Also for each monomial \( S \) define \( |S| \) to be the number of its inversions, the pairs \( k < \ell \) with \( i_k > i_\ell \). To each non-zero tensor \( T = \sum c_S S \) assign a vector \( \rho(T) = (\rho_n) \) where

\[
\rho_n = \sum_{\deg S = n, c_S \neq 0} |S|
\]

for \( n > 0 \). For convenience, set \( \rho_n(0) = 0 \) for all \( n \). Thus \( \rho \) is a vector with an infinite number of coordinates in \( \mathbb{N} \), all but a finite number equal to 0. Define the **reduced degree** of \( T \) to be largest \( n \) with \( \rho_n(T) \neq 0 \). A tensor \( T \) is reduced if and only if its reduced degree is 0. Order such vectors:
\( \nu < \rho \) if and only if either (a) the reduced degree of \( \nu \) is less than that of \( \rho \) or (b) they have the same reduced degree \( d \) but \( \rho_d(\nu) < \rho_d(\rho) \).

Any descending chain of such vectors must be finite. Write \( S \prec T \) if \( \rho(S) < \rho(T) \).

If a tensor \( T \) has a non-reduced term \( A \otimes Y \otimes Z \otimes B \) with \( Y > Z \), it is equivalent modulo \( I \) to

\[
S = A \otimes Z \otimes Y \otimes B + A \otimes [Y, Z] \otimes B
\]

with \( S \prec T \). This can be continued only for a finite number of steps, and at the end we wind up with a reduced tensor. Thus the reduced tensors span the universal enveloping algebra. The hard (and interesting) part is to show that the images of reduced monomials in the enveloping algebra are linearly independent.

I follow the proof to be found in [Bergman:1978]. Bergman’s argument is not all that different from, but somewhat clearer than, the original one of [Birkhoff:1937].

The proof will specify the linear map \( \text{red} \) from \( \bigotimes g \) to the reduced tensors with these properties:

(a) \( T - \text{red}(T) \) lies in \( I \);
(b) \( \text{red}(T) = T \) if all the monomials occurring in \( T \) are ordered;
(c) \( \text{red}(T) = 0 \) if \( T \) lies in \( I \).

This will suffice to prove the Theorem. First of all, since \( T - \text{red}(T) \) lies in \( I \) the image of \( T \) and \( \text{red}(T) \) in the enveloping algebra are the same. Second, if \( T \) is equivalent to two reduced tensors \( T_1 \) and \( T_2 \) then their difference lies in \( I \). Hence \( \text{red}(T_1 - T_2) = 0 \) but then

\[
\text{red}(T_1 - T_2) = 0 = \text{red}(T_1) - \text{red}(T_2) = T_2 - T_2.
\]

Define \( \sigma_{k,\ell} \) to be the linear operator defined on basis elements of the tensor algebra by the formulas

\[
\sigma_{k,\ell}(T) = T \\
\sigma_{k,\ell}(A \otimes X \otimes Y \otimes B) = A \otimes X \otimes Y \otimes B \\
\quad \text{if } \deg A = k, \deg B = \ell, X \leq Y \\
\quad = A \otimes Y \otimes X \otimes B + A \otimes [X, Y] \otimes B \\
\quad \text{if } \deg A = k, \deg B = \ell, X > Y.
\]

Thus \( \sigma_{k,\ell}T - T \) always lies in \( I \), and \( \sigma_{k,\ell}T \prec T \). Write \( S \rightarrow T \) if \( T \) is obtained from \( S \) by a single \( \sigma_{k,\ell} \), and \( S \leadsto T \) if it is obtained from \( S \) by zero or more such operations.

All reductions \( \sigma \) leave irreducible expressions unchanged. If one takes \( B_1 \) and \( B_2 \) in the following Lemma to be irreducible, its conclusion is that \( B_1 = B_2 \).

15.3. Lemma. (PBW Confluence) If \( A \) is a monomial with two reductions \( A \leadsto B_1 \) and \( A \leadsto B_2 \) then there exist reductions \( B_1 \leadsto C \) and \( B_2 \leadsto C \).

Proof. As the following ‘diamond’ diagram suggests, the case of a simple reduction applied several times will prove the general case.
So we must show that if $A$ is a monomial with two one-step reductions $\sigma: A \rightarrow B_1$ and $\tau: A \rightarrow B_2$ then there exist reductions $\sigma': B_1 \rightarrow C$ and $\tau': B_2 \rightarrow C$.

If the reductions are applied to non-overlapping pairs there is no problem. An overlap occurs for a term $A \otimes X \otimes Y \otimes Z \otimes B$ with $X > Y > Z$, say with $|A| = k, |B| = \ell$. It gives rise to a choice of reductions:

$$X \otimes Y \otimes Z \xrightarrow{\sigma_{k,\ell}^{k+1}} Y \otimes X \otimes Z + [X,Y] \otimes Z$$
$$X \otimes Y \otimes Z \xrightarrow{\sigma_{k+1,\ell}^{k+1}} X \otimes Z \otimes Y + X \otimes [Y,Z].$$

But then

$$Y \otimes X \otimes Z + [X,Y] \otimes Z \xrightarrow{\sigma_{k+1,\ell}^{k+1}} Y \otimes Z \otimes X + Y \otimes [X,Z] + [X,Y] \otimes Z$$
$$X \otimes Z \otimes Y + X \otimes [Y,Z] \xrightarrow{\sigma_{k,\ell}^{k+1}} Z \otimes X \otimes Y + [X,Z] \otimes X + Y \otimes [X,Z]$$
and the difference

$$[[Y,Z], X] + [Y,[X,Z]] + [[X,Y], Z]$$

between the right hand sides, because of Jacobi’s identity, lies in $I$.

I call a tensor **uniquely reducible** if there exists exactly one reduced tensor it can be reduced to. We now know that monomials are uniquely reducible. If $S$ is any uniquely reducible element of $\bigotimes g$, let red$(S)$ the unique irreducible element it reduces to.

**15.4. Lemma.** If $S$ and $T$ are uniquely reducible, so is $S + T$, and red$(S + T) = \text{red}(S) + \text{red}(T)$.

**Proof.** Suppose $\sigma$ is a reduction taking $S + T$ to an irreducible expression $W$. According to the previous lemma, we can find a reduction $\sigma'$ such that $\sigma'(\sigma(S)) = \text{red}(S)$. Since $W$ is irreducible, on the one hand we have

$$\sigma'(\sigma(S + T)) = \sigma'(W) = W$$

but on the other it is

$$\sigma'(\sigma(S)) + \sigma'(\sigma(T)) = \text{red}(S) + \sigma'(\sigma(T))$$

Again according to the Lemma, we can find $\sigma''$ such that $\sigma'' \sigma' \sigma(T) = \text{red}(T)$. Then

$$w = \sigma''(W) = \sigma''(\text{red}(S)) + \sigma''(\sigma'(\sigma(T))) = \text{red}(S) + \text{red}(T).$$
An induction argument now implies that every $T$ in $\bigotimes^g$ is uniquely reducible. Define $\text{red}(T)$ to be what it reduces to.

Implicit here is what is called ‘confluence’ in the literature, a tool of great power in finding normal forms for algebraic expressions. It is part of the theory of term rewriting and critical pairs, and although it has been used informally for a very long time, the complete theory seems to have originated in [Knuth-Bendix:1965]. It has been rediscovered independently a number of times. A fairly complete bibliography as well as some discussion of the history can be found in [Bergman:1978].

The ring $U(g)$ is filtered by order. Let $U_n(g)$ be the linear combinations of at most $n$ products of elements of $g$. This filtration is compatible with products, hence determines the graded ring

$$\text{Gr}^n U(g) = \bigoplus_n U_n(g)/U_{n-1}(g).$$

The symbol of a product of $n$ elements in $\bigotimes^g$ the corresponding product in the degree $n$ component $S^n(g)$ of the symmetric algebra of $g$. The kernel contains $U_{n-1}(g)$, and therefore there exists a canonical map

$$\text{Gr}^n U(g) \longrightarrow S^n(g).$$

15.5. Proposition. This canonical map is an isomorphism of graded rings.

15.6. Corollary. If $g$ is the Lie algebra of the Lie group $G$, then the map from $U(g)$ to the ring of all left-invariant differential operators on $G$ generated by the vector fields in $g$ mapping

$$X_1 \otimes X_2 \otimes \ldots \otimes X_n \longmapsto R_{X_1}R_{X_2} \ldots R_{X_n}$$

is an isomorphism.

In other words, the ring $U(g)$ may be described concretely in this case.

Part III. Semi-simple Lie algebras

16. Casimir elements

Suppose now $K$ to be any invariant and non-degenerate inner product on the Lie algebra $g$, for example the Killing form $K_\rho$ associated to a suitable representation $\rho$.

In the following Proposition, let $(X_i)$ be a basis of $g$ and $(X_i^\vee)$ the basis dual with respect to $K$.

16.1. Proposition. The element

$$C_K := \sum X_i X_i^\vee$$

lies in the centre of $U(g)$.

It is called the Casimir element associated to $K$. In the literature one can often find a factor of 2 in this definition, placed there in order to match the Casimir to the Laplacian on the symmetric space attached to the group. If $K = K_{ad}$, this element is just called the Casimir operator $C$ without reference to $K$.

Proof. A straightforward calculation would do, but there is a better way. The bilinear form $K$ gives rise to a covariant linear map $\tau_K$ from $g$ to its linear dual $g^\vee$, and $\tau_K$ is an isomorphism because of non-degeneracy. This in turn gives rise to a $g$-covariant isomorphism of $g \otimes g$ with $g^\vee \otimes g = \text{End}(g)$.

The linear transformation of $g$

$$Y \longmapsto \sum_i \langle X_i^\vee, Y \rangle X_i$$
is the identity transformation, since it takes each \( X_i \) to itself. We have a sequence of maps

\[
\text{Hom}_\mathbb{C}(g, g) \cong g^* \otimes g \xrightarrow{\text{Killing}} g \otimes g \rightarrow U(g)
\]

which are all \( g \)-covariant. The Casimir element \( C_K \) is thus intrinsically characterized as the image of the identity transformation \( I \) on the left. Since \( I \) commutes with \( g \) the Casimir does too, and since \( U(g) \) is generated by \( g \) it lies in the center of \( U(g) \).

In the rest of this section, we’ll look at the case \( g = \mathfrak{sl}_2 \).

The Lie algebra \( g = \mathfrak{sl}_2(F) \) has

\[
\begin{align*}
  h &= \begin{bmatrix} 1 & . \\ . & -1 \end{bmatrix} \\
  e_+ &= \begin{bmatrix} . & 1 \\ . & . \end{bmatrix} \\
  e_- &= \begin{bmatrix} . & . \\ -1 & . \end{bmatrix}
\end{align*}
\]

as a basis. The \( e_{\pm} \) are eigenvectors of \( \text{ad}_h \). The complete specification of the Lie bracket is:

\[
[h, e_{\pm}] = \pm 2e_{\pm},
\]

\[
[e_+, e_-] = -h.
\]

Proposition 13.4 tells us:

**16.2. Corollary.** For the basis \( h, e_{\pm} \) of \( \mathfrak{sl}_2 \) the matrix of the Killing form is

\[
\begin{bmatrix}
8 & 0 & 0 \\
0 & 0 & -4 \\
0 & -4 & 0
\end{bmatrix}
\]

Given the explicit calculation of the form, the Casimir element can be seen to be

\[
C = \frac{1}{8} h^2 - \frac{1}{4} e_+ e_- - \frac{1}{4} e_- e_+,
\]

with alternate expressions

\[
C = h^2/8 - h/4 - e_+ e_-/2 = h^2/8 + h/4 - e_- e_+/2.
\]

Now take \( F \) to be \( \mathbb{R} \), \( G = \text{SL}_2(\mathbb{R}) \), \( K = \text{SO}(2) \subset G \). One reason that the Casimir element is important is because it is related to the Laplacian operator on the Riemannian space \( \mathcal{H} = G/K \). Explicitly:

**16.3. Lemma.** Acting on functions on \( \mathcal{H} \), the Casimir is the same as half the Laplacian.

**Proof.** We know that twice the Casimir is

\[
2C = h^2/4 - h/2 - e_- e_+,
\]

so its action on functions on \( \mathcal{H} \) is as

\[
L_{2C} = (\Lambda_h)^2/4 - \Lambda_h/2 + L_{e_-} L_{e_+}
\]

\[
= \left( x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right)^2 - \left( x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right) - \left( x^2 - y^2 \right) \frac{\partial^2}{\partial x^2} - 2xy \frac{\partial^2}{\partial x \partial y}
\]

\[
= x^2 \frac{\partial^2}{\partial x^2} + 2xy \frac{\partial^2}{\partial x \partial y} + y^2 \frac{\partial^2}{\partial y^2} - (x^2 - y^2) \frac{\partial^2}{\partial x^2} - 2xy \frac{\partial^2}{\partial x \partial y}
\]

\[
= y^2 \frac{\partial^2}{\partial x^2} + y^2 \frac{\partial^2}{\partial y^2}
\]

\[
= \Delta_{\mathcal{H}}.
\]
17. Semi-simple Lie algebras

If \( \mathfrak{h} \) and \( \mathfrak{t} \) are solvable ideals of \( \mathfrak{g} \) then so is \( \mathfrak{h} + \mathfrak{t} \). Therefore there exists a maximal solvable ideal, which I shall call the Lie radical of \( \mathfrak{g} \). A Lie algebra is called semi-simple if its Lie radical is 0. The Killing radical of \( \mathfrak{g} \) is the radical of the Killing form \( K_{ad} \)—i.e. the subspace of all \( X \) in \( \mathfrak{g} \) such that \( K(X, g) = 0 \).

17.1. Lemma. The Killing radical is a solvable ideal of \( \mathfrak{g} \).

Proof. Let \( \mathfrak{r} \) be the Killing radical of \( \mathfrak{g} \). It is immediate from the invariance of the Killing form that \( \mathfrak{r} \) is an ideal of \( \mathfrak{g} \). If \( x \) lies in \( \mathfrak{r} \) then \( K(x, y) = 0 \) for all \( y \) in \( \mathfrak{g} \), hence in particular for \( y \) in \( D\mathfrak{r} \). Hence by Cartan’s criterion \( ad_{\mathfrak{r}} \mathfrak{g} \) is solvable. But \( \mathfrak{r} \) is an extension of this by some subspace of the centre of \( \mathfrak{g} \), so that \( \mathfrak{r} \) itself must be solvable.

17.2. Theorem. A Lie algebra is semi-simple if and only if its Killing form \( K_{ad} \) is non-degenerate.

Proof. If the Killing form is degenerate then the Killing radical is non-trivial, and by the Lemma \( \mathfrak{g} \) contains a non-trivial solvable ideal, hence cannot be semi-simple.

For the other half, suppose that the Lie radical \( \mathfrak{r} \) is not 0. Then either \( \mathfrak{r} \) is abelian, or \( D\mathfrak{r} \) is nilpotent and not 0. In the latter case, the centre of \( D\mathfrak{r} \) is an abelian ideal of \( \mathfrak{g} \). In either case, we may assume that \( \mathfrak{g} \) contains a non-trivial abelian ideal \( \mathfrak{a} \). Then \( K(a, x) = 0 \) for all \( a \) in \( \mathfrak{a} \), \( x \) in \( \mathfrak{g} \), so \( \mathfrak{a} \) is contained in the Killing radical of \( \mathfrak{g} \).

17.3. Theorem. If \( \mathfrak{g} \) is semi-simple and \( \mathfrak{h} \) is an ideal of \( \mathfrak{g} \), then \( \mathfrak{h} \perp \) is a complementary ideal.

Proof. Because, as the argument just finished shows, \( \mathfrak{h} \cap \mathfrak{h} \perp \) is solvable.

This can be improved somewhat:

17.4. Proposition. If \( \mathfrak{g} \) is a semi-simple ideal in a Lie algebra \( \mathfrak{h} \), there exists a complementary ideal in \( \mathfrak{h} \).

Proof. Consider \( \mathfrak{h} \) as a module over \( \mathfrak{g} \), via the adjoint representation. Let \( \mathfrak{a} \) be a \( \mathfrak{g} \)-stable complement to \( \mathfrak{g} \) itself. It is easy to see that \( [\mathfrak{g}, \mathfrak{a}] = 0 \), and in fact \( \mathfrak{a} \) is the space of all \( y \) such that \( [\mathfrak{g}, y] = 0 \). Therefore \( \mathfrak{a} \) is uniquely characterized, and it is an ideal in \( \mathfrak{h} \) because it is the annihilator of the \( \mathfrak{h} \)-ideal \( \mathfrak{g} \).

A simple Lie algebra is one with no non-trivial ideals. If it has dimension greater than one, it will be semi-simple, in which case I’ll call it a proper simple Lie algebra.

17.5. Corollary. Any semi-simple Lie algebra is the direct sum of simple Lie algebras.

17.6. Corollary. If \( \mathfrak{g} \) is semi-simple then \( \mathfrak{g} = D\mathfrak{g} \).

Proof. Because it is clearly true of a simple algebra.

The following is the principal result of this section, and motivates the term “semi-simple.”

17.7. Theorem. Any finite-dimensional representation of a semi-simple Lie algebra decomposes into a direct sum of irreducible representations.

Proof. I follow LA §6.3 of [Serre:1965], which was perhaps the first clear account of the purely algebraic proof (as opposed to the one of Hermann Weyl for complex groups, which utilized the relationship between complex semi-simple groups and their maximal compact subgroups).

The proof is by induction on dimension. If \( V = \{0\} \) there is no problem. Otherwise we can find a proper \( \mathfrak{g} \)-stable subspace \( U \) and a short exact sequence of \( \mathfrak{g} \)-representations

\[
0 \to U \to V \to W \to 0
\]

where \( W \) is irreducible. By induction, \( U \) decomposes into a direct sum of irreducibles, so the Theorem will follow from this:

17.8. Lemma. An exact sequence of \( \mathfrak{g} \)-modules as above, with \( W \) irreducible, splits.
Proof of the Lemma. We assume at first \( W \) to be the trivial representation \( \mathbb{C} \), so our exact sequence is

\[
0 \to U \to V \to \mathbb{C} \to 0
\]

We proceed now by induction on the dimension of \( U \). If it is 0, there is nothing to prove. Otherwise, \( U \) will contain an irreducible \( g \) subspace \( U' \).

Case 1. The subspace \( U' \) is not all of \( U \). We have an exact sequence

\[
0 \to U/U' \to V/U' \to \mathbb{C} \to 0
\]

which splits by the induction assumption. Therefore \( V/U' \) contains a one-dimensional \( g \)-stable subspace, so we can write

\[
V/U' = U/U' \oplus \mathbb{C}.
\]

If \( V' \) is the inverse image in \( V \) of \( \mathbb{C} \), we have an exact sequence

\[
0 \to U' \to V' \to \mathbb{C} \to 0
\]

which again splits by induction, giving a one-dimensional subspace of the original \( V \).

Case 2. The subspace \( U \) in the sequence is irreducible.

There is a further subdivision. (a) Suppose \( U \) is trivial. We are looking at this situation:

\[
0 \to F \to V \to F \to 0.
\]

If \( X \) and \( Y \) are elements of \( g \) then \( XY \) and \( YX \) are both 0 on \( V \), hence \([X, Y] = 0\) as well. But by Corollary 17.6 \( g = \mathfrak{d}g \) so all of \( g \) acts trivially on \( V \), which must be \( F \oplus F \).

(b) The representation \( \rho \) on \( U \) is not trivial. The image \( g_\rho \) of \( g \) in \( \text{End}(U) \) factors through a direct sum of simple algebras, and the Killing form \( K_\rho \) also factors through \( g_\rho \).

The Killing for \( K_\rho \) is non-degenerate on \( g_\rho \).

Similar things have been proved before, I leave this claim as an exercise.

Because \( \rho \) is irreducible, the Casimir \( C_\rho \) acts as a scalar, say \( \gamma_\rho \). Its trace is on the one hand \( \gamma_\rho \text{dim} V \), and on the other

\[
\text{trace} C_\rho = (1/2) \sum_i \text{trace}(\rho(X_i)\rho(X_i^\vee)) = \frac{\text{dim} g}{2},
\]

so \( \gamma_\rho \neq 0 \). But \( C_\rho \) lies in the universal enveloping algebra of \( g \), and acts as 0 on \( F \). So \( V \) decomposes into a direct sum of eigenspaces with respect to \( C_\rho \).

At this point, we know that any sequence

\[
0 \to U \to V \to F \to 0
\]

splits.

Suppose now that we have an arbitrary exact sequence of \( g \)-spaces

\[
0 \to U \to V \to W \to 0
\]

with \( W \) irreducible, and consider the exact sequence of \( g \)-modules

\[
0 \to \text{Hom}_F(W, U) \to \text{Hom}_F(W, V) \to \text{Hom}_F(W, W) \to 0.
\]

Let \( V' \) be the subspace of maps in \( \text{Hom}_F(W, V) \) mapping onto scalar multiplications. Then we have a sequence

\[
0 \to \text{Hom}_F(W, U) \to V' \to F \to 0
\]

But since this splits as a \( g \)-module, there exists a \( g \)-invariant element in \( V' \subseteq \text{Hom}_F(W, V) \) mapping onto the identity map from \( W \) to itself. This amounts to a splitting. 

\[\square\]
18. Representations of SL(2)

The Lie algebra $\mathfrak{g} = \mathfrak{sl}(2, F)$ has

$$h = \begin{bmatrix} 1 & \cdot \\ \cdot & -1 \end{bmatrix},$$
$$e_+ = \begin{bmatrix} \cdot & 1 \\ \cdot & \cdot \end{bmatrix},$$
$$e_- = \begin{bmatrix} \cdot & \cdot \\ -1 & \cdot \end{bmatrix},$$

as a basis. The $e_{\pm}$ are eigenvectors of $\text{ad}_h$. The complete specification of the Lie bracket is:

$$[h, e_{\pm}] = \pm 2e_{\pm}$$
$$[e_+, e_-] = -h.$$

Suppose $V$ to be an irreducible finite-dimensional module over $\mathfrak{g}$. If $\overline{V}$ is $V \otimes_F F$ (where $F$ is an algebraic closure of $F$) there exists at least one eigenvalue of $h$:

$$h \cdot v = \lambda v.$$

Then $e_{\pm} \cdot v$ is also an eigenvector with eigenvalue $\lambda \pm 2$, and likewise $e_{k \pm} \cdot v_0$ is $g$-stable. It must therefore be all of $\overline{V}$. The vector $v_0$ is an eigenvector of $h$ with eigenvalue $\mu = 2\ell$. Finite-dimensionality implies that $e_- \cdot v_0 = 0$ for some $n$.

The Casimir element $C$ of $U(\mathfrak{g})$ acts as a scalar on $\overline{V}$. Since

$$C = h^2/4 + h/2 + e_- e_+ - h^2/4 + h/2 + e_+ e_-,$$

we have

$$C v_0 = (\mu^2/4 + \mu/2) v_0$$
$$C u_0 = ((\mu - 2n)^2/4 - (\mu - 2n)/2) u_0$$
$$\mu^2/4 + \mu/2 = \mu^2 - 4\mu n + 4n^2 - \mu - 2n/2$$
$$0 = -\mu n + 4n^2 - \mu + n$$
$$(n + 1)(\mu - n) = 0$$
$$n = \mu.$$

Consequently, $\mu$ must be a non-negative integer. Hence there exist eigenvectors already in the original $F$-space $V$, and one may (and I shall) assume all the $v_\ell$ to be in $V$ itself.

18.1. Proposition. Suppose $V$ to be a vector space over $F$, and an irreducible module over $\mathfrak{sl}_2(F)$. There exists a vector $v_0$ with $e_+ \cdot v_0 = 0$ and $h \cdot v_0 = \mu v_0$ for some integer $\mu \geq 0$. The space $V$ is then spanned by the vectors $v_\ell = e_\ell \cdot v_0$ for $0 \leq \ell \leq n$ with

$$h \cdot v_\ell = (n - 2\ell)v_\ell, \quad e_- \cdot v_\ell = 0.$$
18.2. Lemma. If $e_v \cdot v_0 = 0$ and $h \cdot v_0 = \lambda v_0$ then

\[ e_+ e^- v_0 = k(k - 1 - \lambda)e^{-1}_- v_0. \]

Proof. Say $e_+ e^- v_0 = \lambda_k e^{-1}_- v_0$. Then $\lambda_{k+1} = \lambda_k - \lambda + 2k$, and $\lambda_0 = 0$. The formula above follows by induction.

It can be verified directly by algebraic calculation that the formulas above define an irreducible representation of dimension $n + 1$ of $\mathfrak{sl}_2$. But these representations can be constructed explicitly. Any finite-dimensional representation of $\text{SL}_2(F)$ gives rise to a representation of its Lie algebra $\mathfrak{sl}_2(F)$. There are two obvious representations—the trivial one and the tautological representation $\pi_1$ on $F^2$. The associated representation of $\mathfrak{sl}_2$ acts like this on the standard basis $u = [1, 0], v = [0, 1]$:

- $h : u \mapsto u$
- $v \mapsto -v$
- $e_+ : v \mapsto u$
- $u \mapsto 0$
- $e_- : u \mapsto -v$
- $v \mapsto 0$

which can be illustrated (but not indicating a necessary $\rightarrow$-sign):

```
\[
\begin{array}{c c c c c c}
\ & \ & \ & \ & \ & \\
\ & v + & & & & \\
\ & \ & \ & & & \\
\ & \ & \ & & & \\
\ & \ & \ & & \ & \\
\ & \ & \ & & \ & \\
\ & \ & \ & & \ & \\
\end{array}
\]
```

This representation of $\text{SL}_2(F)$ on $V = F^2$ gives rise to the representation $\pi_n$ on the symmetric product $S^n V$ with basis $u^k v^{n-k}$ for $0 \leq k \leq n$:

\[ \pi_n(g) : u^k v^{n-k} \mapsto (gu)^k (gv)^{n-k}. \]

Thus

\[ \pi_n \begin{pmatrix} 1 & x \\ - & 1 \end{pmatrix} : u^k v^{n-k} \mapsto u^k (v + xu)^{n-k} \]

which implies

\[ \pi_n(e_+) : u^k v^{n-k} \mapsto (n-k)u^{k+1} v^{n-k-1}. \]

Similarly

\[ \pi_n(e_-) : u^k v^{n-k} \mapsto (-1)^k ku^{k-1} v^{n-k+1}. \]

If $F = \mathbb{C}$ we can calculate $\pi_n(h)$ by applying the exponential and differentiating. In general, we can proceed formally by using the nil-ring $F[e]$ with $e^2 = 0$ or by using the formula $[e_+, e_-] = -h$ to deduce

\[ \pi_n(h) : u^k v^{n-k} \mapsto -(n-2k)u^k v^{n-k}. \]

Also

\[ e_- e_+ : u^k v^{n-k} \mapsto (-1)^{k+1}(k+1)(n-k)u^k v^{n-k} \]

\[ e_+ e_- : u^k v^{n-k} \mapsto (-1)^k k(n-k+1)u^k v^{n-k}. \]

For example, when $n = 3$:
There is one more simple formula that is useful. If
\[ w = \begin{bmatrix} -1 & 1 \\ -1 & \cdot \end{bmatrix} \]
then
\[ \pi_n(w) : u^\ell v^{n-\ell} \mapsto (1)^\ell u^{n-\ell} v^\ell, \]
and if \( n = 2\ell + 1 \)
\[ \pi_n(w) : u^{\ell+1} v^{\ell+1} \mapsto (-1)^\ell u^{\ell+1} v^{\ell+1} \]
\[ \pi_n(e_+) : u^{\ell+1} v^{\ell+1} \mapsto (-1)^\ell (\ell + 1) \pi_n(w) u^\ell v^{\ell+1}. \]

Every finite-dimensional representation of \( \mathfrak{sl}_2 \) is a direct sum of irreducible ones, and the irreducible ones are representations of \( SL_2(F) \). Hence:

18.3. Proposition. Every finite-dimensional representation of \( \mathfrak{sl}_2 \) is derived from a representation of the group \( SL_2 \).

This is the principal consequence of this section in what follows. But another consequence is this:

18.4. Proposition. If \( V \) is a finite-dimensional module over \( \mathfrak{sl}_2 \), and \( v \) in \( V \) is an eigenvector for \( h \) annihilated by \( e_+ \), its eigenvalue with respect to \( h \) is non-negative.

19. Tensor invariants

I learned the following result from LA §6.5 of [Serre:1966].

If we are given an embedding of \( \mathfrak{g} \) into \( gl(V) \), then there are associated representations of \( \mathfrak{g} \) on the dual \( \hat{V} \) and on the tensor products
\[ \otimes^{p,q} V = \otimes^p V \otimes \otimes^q \hat{V}. \]

A tensor invariant for \( \mathfrak{g} \) is any tensor annihilated by it. One example is the identity map in \( \text{End}(V) \), which may be identified with \( \hat{V} \otimes V \).

19.1. Theorem. If \( \mathfrak{g} \subseteq gl(V) \) is semi-simple, it is same as the subalgebra of \( gl(V) \) leaving invariant all the tensor invariants of \( \mathfrak{g} \).

Proof. Let \( \mathfrak{h} \) be the Lie algebra annihilating all the tensor invariants of \( \mathfrak{g} \). We know that \( \mathfrak{g} \subseteq \mathfrak{h} \), and we must show equality. This comes from a series of elementary steps interpreting tensors and tensor invariants suitably. For details, look at LA §6.5 of Serre’s book.

(a) Any \( \mathfrak{g} \)-homomorphism from \( \otimes^{p,q} V \) to \( \otimes^{r,s} V \) is also an \( \mathfrak{h} \)-homomorphism.

(b) Any \( \mathfrak{g} \)-stable subspace of \( \otimes^{p,q} V \) is also \( \mathfrak{h} \)-stable.

(c) By Proposition 17.4, we may write \( \mathfrak{h} = \mathfrak{g} \oplus \mathfrak{c} \), where \( \mathfrak{c} \) is an \( \mathfrak{h} \)-ideal commuting with \( \mathfrak{g} \). But then by (a) it also commutes with \( \mathfrak{h} \). This means that \( \mathfrak{c} \) is the centre of \( \mathfrak{h} \).

(d) If \( U \) is an irreducible \( \mathfrak{g} \)-submodule of \( V \), then \( \mathfrak{c} \) acts by scalars on it. In fact, it acts by 0. Why? The exterior products \( \wedge^k W \) are contained in the tensor algebra. The semi-simple algebra \( \mathfrak{g} \) acts trivially on its highest exterior product, and therefore so does \( \mathfrak{h} \), hence \( \mathfrak{c} \). But the action of any element \( X \) of \( gl(V) \) on this power is by the scalar trace(\( X \)), which vanishes if \( X \) lies in \( \mathfrak{c} \).

If \( T \) is any linear transformation in \( M_n(F) \), where \( F \) is a field of characteristic 0, it may be written uniquely as \( T_s + T_n \), where \( T_s \in M_n(F) \) is diagonalizable over an algebraic closure, \( T_n \in \text{nil} M_n(F) \) is nilpotent, and
Proof\textsuperscript{a}. I recall that this means that $K$ is orthogonal to $a$, hence must be a degenerate on $\mathcal{D}$, since $K$ is orthogonal. The restriction of $\pi$ to any summand must be non-degenerate.

20. The structure of semi-simple Lie algebras

In this section, let $\mathfrak{g}$ be a semi-simple Lie algebra, and $\mathfrak{a}$ a Cartan subalgebra, which exists by Theorem 11.1. I recall that this means that $\mathfrak{a}$ is a nilpotent algebra that is its own normalizer in $\mathfrak{g}$. By Theorem 10.8, the adjoint action of $\mathfrak{a}$ on $\mathfrak{g}$ decomposes into a direct sum of primary modules $\mathfrak{g}_\lambda$. These are called root spaces when $\lambda \neq 0$.

Two spaces $\mathfrak{g}_\lambda$ and $\mathfrak{g}_\mu$ are orthogonal with respect to the Killing form $K = K_{\text{ad}}$, so the decomposition

$$\mathfrak{g} = \mathfrak{a} + \bigoplus_{\lambda \neq 0} (\mathfrak{g}_\lambda \oplus \mathfrak{g}_{-\lambda})$$

is orthogonal. The restriction of $K$ to any summand must be non-degenerate.

20.1. Proposition. Every Cartan subalgebra of $\mathfrak{g}$ is abelian and is its own centralizer.

Proof. According to Cartan’s criterion, $\text{trace}(\text{ad}_X, \text{ad}_Y) = 0$ for every $X$ in $\mathfrak{a}$, $Y$ in $\mathcal{D}\mathfrak{a}$. Since $K$ is non-degenerate on $\mathfrak{a}$, we must have $\mathcal{D}(\mathfrak{a}) = 0$. Thus the centralizer of $\mathfrak{a}$ contains $\mathfrak{a}$ and is contained in its normalizer, hence must be $\mathfrak{a}$.

20.2. Proposition. Every element of $\mathfrak{a}$ is semi-simple.

Proof. The nilpotent component in the Jordan decomposition, which lies in $\mathfrak{g}$ according to Corollary 19.2 is orthogonal to $\mathfrak{a}$.

For $\lambda$ a root of $\mathfrak{g}$ with respect to $\mathfrak{a}$, let $h_\lambda$ be the inverse image of $\lambda$ under the isomorphism $\mathfrak{a} \rightarrow \mathfrak{h}$ induced by $K$.

20.3. Proposition. Suppose $X$ in $\mathfrak{g}_\lambda$, $Y$ in $\mathfrak{g}_\mu$. Then $K(X, Y) = 0$ unless $\mu = -\lambda$, in which case $[X, Y] = K(X, Y) h_\lambda$.

Proof. Only the last assertion is troublesome. For any $H$ in $\mathfrak{a}$

$$K(H, [X, Y]) = K([H, X], Y) = \lambda(H) \cdot K(X, Y) = K(H, h_\lambda) K(X, Y),$$

since $K$ is invariant.

20.4. Proposition. There exists some multiple $H_\lambda$ of $h_\lambda$ such that $\langle \lambda, H_\lambda \rangle = 2$.

Proof. It must be shown that $\langle \lambda, h_\lambda \rangle \neq 0$. Suppose the contrary. Since the restriction of $K_{\text{ad}}$ is non-degenerate, by the previous result we may choose $e$ in $\mathfrak{g}_\lambda$ and $f$ in $\mathfrak{g}_{-\lambda}$ with $h = [e, f] \neq 0$. By assumption $\langle \lambda, h \rangle = 0$, so

$$[h, e] = 0, \quad [h, f] = 0, \quad [x, y] = h.$$

So the Lie algebra generated by them is solvable. Lie’s criterion tells us that $\text{ad}_h$ is nilpotent. But it is also semi-simple, so it must be 0. Contradiction.
If we choose \( e \neq 0 \) in \( \mathfrak{g}_\lambda \), we may find \( f \) in \( \mathfrak{g}_{-\lambda} \) such that \([e, f] = -h = -H_\lambda\). We then have

\[
\begin{align*}
[&h, e] = 2e, \\
[&e, f] = -h, \\
[&h, f] = -2f
\end{align*}
\]

which tells us that the subspace of \( \mathfrak{g} \) spanned by \( e, h, \) and \( f \) is a Lie algebra isomorphic to \( \mathfrak{sl}_2 \). Hence we get the adjoint representation of this copy of \( \mathfrak{sl}_2 \) on \( \mathfrak{g} \). It splits into a sum of irreducible representations, and this lifts to a representation of the group \( \text{SL}_2 \). These copies of \( \mathfrak{sl}_2 \) and the associated representations of \( \text{SL}_2 \) are part of the basic structure of a semi-simple Lie algebra.

20.5. Proposition. The eigenspace \( \mathfrak{g}_\lambda \) of the root \( \lambda \) has dimension one.

Proof. It suffices to prove that every \( \mathfrak{g}_{-\lambda} \) has dimension one. The subspace \( e^\perp \) in \( \mathfrak{g}_{-\lambda} \) has codimension one. Say \( f_* \neq 0 \) lies in \( e^\perp \). By Proposition 20.3, \([e, f_*] = 0\). But then we have a vector in a finite-dimensional module over \( \mathfrak{sl}_2 \) annihilated by \( e \) of weight \(-2\) with respect to \( h \), which contradicts Proposition 18.4.

This means that a choice of \( e_\lambda \neq 0 \) in \( \mathfrak{g}_\lambda \) is unique up to scalar, and then we get \( e_{-\lambda}, H_\lambda \) generating a unique copy of \( \mathfrak{sl}_2 \) in \( \mathfrak{g} \) that I’ll call \( \mathfrak{sl}_2,\lambda \). The adjoint representation of this subalgebra on \( \mathfrak{g} \) splits up into a direct sum of irreducible representations, and is derived from a direct sum of representations of the associated group \( \text{SL}_2,\lambda \). Let \( s_\lambda \) be the involutory automorphism of \( \mathfrak{g} \) corresponding to the image of

\[
\begin{pmatrix}
0 & -1 \\
1 & 0
\end{pmatrix}
\]

in \( \mathfrak{g} \). It commutes with the kernel of \( \lambda \) in \( \mathfrak{a} \), which is a complement to \( H_\lambda \). This involution take \( H_\lambda \) to \(-H_\lambda\), so normalizes \( \mathfrak{a} \). Since it acts as an automorphims on all of \( \mathfrak{g} \), this proves:

20.6. Proposition. The reflection \( s_\alpha \) takes \( \Sigma \) to itself.

20.7. Proposition. The set \( \Sigma \) spans \( \hat{\mathfrak{h}} \).

Proof. An element \( h \) in the complement acts trivially on \( \mathfrak{g} \), hence lies in the centre of \( \mathfrak{g} \).

The involutions \( s_\lambda \) also preserve the Killing form.

According to [Bourbaki:1972] a root system is a subset \( \Sigma \) of a real vector space \( V \) satisfying these conditions:

(a) \( \Sigma \) is finite, does not contain 0, and generates \( V \);
(b) for each \( \lambda \) in \( \Sigma \) there exists \( \lambda^\vee \) in the dual of \( V \) such that \( \langle \lambda, \lambda^\vee \rangle = 2 \) and the reflection

\[
s_\lambda: v \mapsto v - \langle v, \lambda^\vee \rangle \lambda
\]

takes \( \Sigma \) to itself;
(c) for all \( \lambda \) in \( \Sigma \), \( \langle \lambda, \mu \rangle \in \mathbb{Z} \).

I have verified all of these for our \( \Sigma \) except (c). This follows from facts about finite-dimensional representations of \( \text{SL}_2 \). Therefore:

20.8. Proposition. The set \( \Sigma \) is a root system in \( \hat{\mathfrak{h}} \).

But what this really means, and what its consequences are, I do not consider here.
Part IV. References


