Lie Theory, Universal Enveloping Algebras, and the Poincaré-Birkhoff-Witt Theorem

Lucas Lingle

August 22, 2012

Abstract

We investigate the fundamental ideas behind Lie groups, Lie algebras, and universal enveloping algebras. In particular, we emphasize the useful properties of the exponential mapping, which allows us to transition between Lie groups and Lie algebras. From there, we discuss universal enveloping algebras, prove their existence and uniqueness, and after introducing the necessary machinery, we prove the Poincaré-Birkhoff-Witt Theorem.

1 Introduction

In the first section, we introduce Lie groups and prove some basic theorems about them. In the second section, we discuss and prove the properties of the exponential mapping. In the third section, we introduce Lie algebras and prove some important facts relating Lie groups to Lie algebras. In the fourth section, we introduce universal enveloping algebras, and prove their existence and uniqueness. In the fifth and final section, we prove the Poincaré-Birkhoff-Witt Theorem and its corollaries.

2 Lie Groups

Definition 2.1. A *Lie group* G is group which is also a finite-dimensional smooth manifold, and in which the group operation and inversion are smooth maps.

Definition 2.2. The *general linear group* over the real numbers, denoted $GL_n(\mathbb{R})$, is the set of all $n \times n$ invertible real matrices, equipped with the operation of matrix multiplication.

Similarly, the *general linear group* over the complex numbers, denoted $GL_n(\mathbb{C})$, is the set of all $n \times n$ invertible complex matrices, equipped with the operation of matrix multiplication.

Since the general linear groups only contain invertible matrices, each matrix in $GL_n(\mathbb{R})$ has an inverse in $GL_n(\mathbb{R})$, so the general linear groups are closed under inversion. Since the product AB of any two invertible matrices A and B is also invertible, and has entries in the same field as A and B, the general linear groups are closed under the group operation. Lastly, since matrix multiplication is associative, the elements of $GL_n(\mathbb{R})$ associate. Hence, $GL_n(\mathbb{R})$ is a group. The above logic likewise holds for $GL_n(\mathbb{C})$.

More abstractly, the general linear group of a vector space V, written $GL(V)$, is the automorphism group, whose elements can be written in matrix form but can also be thought of as operators that form a group under composition.

Definition 2.3. Denote the set of all $n \times n$ complex matrices by $M_n(\mathbb{C})$.

Definition 2.4. Let $\{A_m\}$ be a sequence of complex matrices in $M_n(\mathbb{C})$. We say that $\{A_m\}$ *converges to a matrix* A if each entry of the matrices in the sequence converges to the corresponding entry in A. That is, if $(A_m)_{kl}$ converges to A_{kl} for all $1 \leq k, l \leq n$, we say $\{A_m\}$ converges to A.

Definition 2.5. A *matrix Lie group* is any subgroup G of $GL_n(\mathbb{C})$ such that if $\{A_m\}$ is any sequence of matrices in G converging to some matrix A, then either A is in G or else A is not invertible.

Thus a matrix Lie group is a set algebraically closed under the inherited group operation from $GL_n(\mathbb{C})$, and is also a topologically closed subset of $GL_n(\mathbb{C})$. In other words, a matrix Lie group is a closed subgroup of $GL_n(\mathbb{C})$.

Definition 2.6. A matrix Lie group G is said to be *compact* if the following two conditions are satisfied:

1. If $\{A_m\}$ is any sequence of matrices in G and $\{A_m\}$ converges to a matrix A, then A is in G.
2. There is some $C \in \mathbb{R}$ such that for all matrices $A \in G$, $|A_{ij}| \leq C$ for all $1 \leq i, j \leq n$.

Definition 2.7. A matrix Lie group G is *connected* if given any two matrices $A, B \in G$, there exists a continuous path $A(t)$, for $a \leq t \leq b$, so that $A(a) = A$ and $A(b) = B$.

Technically, this is what is known as path-connectedness in topology, and generally is not the same as connectedness. However, a matrix Lie group is connected if and only if it is path connected, and so we shall continue to refer to matrix Lie groups as connected when they are path-connected.

Definition 2.8. A matrix Lie group G that is not connected can be uniquely described as the union of disjoint sets. Each such disjoint set is called a *component* of G.

Proposition 2.9. *If G is a matrix Lie group, then the component of G containing the identity is a subgroup of G.*

Proof. Let A and B be two matrices in the component of G containing the identity. Then there exist two continuous paths $A(t)$ and $B(t)$, with $A(0) = B(0) = I$, $A(1) = A$, and $B(1) = B$. Then $A(t)B(t)$ is a continuous path from I to AB. But A and B are any two elements of the identity component, and their product AB is also in the identity component, since the continuous path given by $A(t)B(t)$ goes from I to AB and such a continuous path can only be formed between elements of the same component. Let $(A(t))^{-1}$ denote the inverse of the matrix given by $A(t)$, for each t. Then $(A(t))^{-1}$ goes from I to A^{-1}, and by the same logic as above, A^{-1} must be in the identity component as well. Since the identity component is closed under the inherited group operation and under inversion, it is a subgroup of G. ■

Definition 2.10. Let G and H be matrix Lie groups. A map $\Phi : G \to H$ is called a *Lie group homomorphism* if Φ is continuous and $\Phi(g_1 g_2) = \Phi(g_1)\Phi(g_2)$ for all $g_1, g_2 \in G$.

If Φ is a bijective Lie group homomorphism and Φ^{-1} is continuous, then Φ is called a *Lie group isomorphism.*

3 The Exponential Mapping

Although Lie groups are endowed with some extra structure and thus are an easier form of manifold to study, they themselves can still be difficult to deal with. For this reason, we often deal with a more wieldy object, namely the Lie algebra corresponding to the group. In order to transfer information from the Lie algebra to the Lie group, we use a function called the exponential mapping.

Definition 3.1. Let X be any matrix. Define the *matrix exponential* by

$$e^X = \sum_{m=0}^{\infty} \frac{X^m}{m!}.$$

One might wonder if this even converges. As we will see shortly, the answer is an emphatic yes. First, though, we must introduce a few new concepts.

Definition 3.2. The *Hilbert-Schmidt norm* of an $n \times n$ matrix X is given by

$$||X|| = \left(\sum_{j=1}^{n} \sum_{i=1}^{n} |x_{ij}|^2 \right)^{1/2}.$$

It is easy to verify, using the triangle and Cauchy-Schwarz inequalities, that the norm obeys the following:

$$||X + Y|| \leq ||X|| + ||Y||,$$

$$||XY|| \leq ||X||\,||Y||.$$

Proposition 3.3. *For any $n \times n$ real or complex matrix X, the series above converges. Furthermore, e^X is a continuous function of X.*

Proof. Since we are working with matrices having real or complex entries, we know that there is some entry whose absolute value is the greatest among the entries. Let M denote the maximum, in absolute value, of all entries of the matrix X. Then $|(X)_{ij}| \leq M$, and since X is a $n \times n$ matrix, $|(X^2)_{ij}| \leq nM^2$, and so on. In general, $|(X^m)_{ij}| \leq n^{m-1}M^m$. Then

$$\sum_{m=0}^{\infty} \frac{n^{m-1}M^m}{m!}$$

converges by a simple application of the ratio test. Then since $|(X^m)_{ij}| \leq n^{m-1}M^m$, we can use the comparison test. Thus, the sum

$$\sum_{m=0}^{\infty} \frac{|(X^m)_{ij}|}{m!} = \sum_{m=0}^{\infty} \left| \left(\frac{X^m}{m!} \right)_{ij} \right|$$

converges as well. Then by a basic theorem from analysis, we know that since

$$\sum_{m=0}^{\infty} \left(\frac{X^m}{m!} \right)_{ij}$$

converges absolutely, it converges in general. By Definition 2.4, we know the sequence (of partial sums) of matrices converges—and hence

$$\sum_{m=0}^{\infty} \frac{X^m}{m!} = e^X$$

converges. It is easy to see that e^X is continuous. ■

Now that we see that the exponential mapping is well-behaved, we can prove some important properties about it.

Proposition 3.4. *Let X and Y be arbitrary $n \times n$ matrices, and let M^* denote the conjugate transpose of a matrix M. Then we have the following:*

1. $e^0 = I$,
2. $(e^X)^* = e^{(X^*)}$,
3. e^X *is invertible and* $(e^X)^{-1} = e^{-X}$,
4. $e^{(\alpha+\beta)X} = e^{\alpha X}e^{\beta X}$ *for all* $\alpha, \beta \in \mathbb{C}$,
5. *if* $XY = YX$, *then* $e^{X+Y} = e^X e^Y = e^Y e^X$,
6. *if* C *is invertible, then* $e^{CXC^{-1}} = Ce^X C^{-1}$,
7. $||e^X|| \leq e^{||X||}$.

Proof. Point 1 is obvious, and Point 2 follows from taking the conjugate transposes term-wise. Points 3 and 4 are special cases of Point 5.

For Point 5, we note that since e^Z converges for all Z, $e^X e^Y$ is defined for all X and Y. Furthermore,

$$e^X e^Y = \left(I + X + \frac{X^2}{2!} + \cdots\right)\left(I + Y + \frac{Y^2}{2!} + \cdots\right).$$

Multiplying out, and collecting terms where the power of X plus the power of Y is m, we get

$$e^X e^Y = \sum_{m=0}^{\infty}\sum_{k=0}^{m} \frac{X^k}{k!}\frac{Y^{m-k}}{(m-k)!} = \sum_{m=0}^{\infty}\frac{1}{m!}\sum_{k=0}^{m}\frac{m!}{k!(m-k)!}X^k Y^{m-k}.$$

And since X and Y commute,

$$(X+Y)^m = \sum_{k=0}^{m}\frac{m!}{k!(m-k)!}X^k Y^{m-k}.$$

So we get

$$e^X e^Y = \sum_{m=0}^{\infty}\frac{1}{m!}(X+Y)^m = e^{(X+Y)}.$$

Point 6 follows immediately, since each term of the matrix exponential can be written as

$$\frac{(CXC^{-1})^m}{m!} = \frac{(CXC^{-1})(CXC^{-1})\cdots(CXC^{-1})(CXC^{-1})}{m!} = C\left(\frac{X^m}{m!}\right)C^{-1}.$$

For Point 7, notice that for each $m \in \mathbb{N}$, by the Cauchy-Schwarz inequality,

$$\left|\left|\frac{X^m}{m!}\right|\right| = \frac{||X^m||}{m!} \leq \frac{||X||^m}{m!}.$$

And since $||X||$ is a real number,

$$e^{||X||} = \sum_{m=0}^{\infty}\frac{||X||^m}{m!}$$

converges. By the comparison test, we know that

$$\sum_{m=0}^{\infty}\left|\left|\frac{X^m}{m!}\right|\right|.$$

converges as well. It follows from the triangle inequality that

$$S_K := \left|\left|\sum_{m=0}^{K}\frac{X^m}{m!}\right|\right| \leq \sum_{m=0}^{K}\left|\left|\frac{X^m}{m!}\right|\right| \leq \sum_{m=0}^{K}\frac{||X||^m}{m!} =: L_K.$$

Since the sequence defined by

$$E_K := \sum_{m=0}^{K} \frac{X^m}{m!}$$

converges (to e^X), we know that the sequence

$$S_K := \left|\left| \sum_{m=0}^{K} \frac{X^m}{m!} \right|\right|$$

converges as well (to $||e^X||$). It follows that

$$\lim_{K\to\infty} S_K = \left|\left|e^X\right|\right| \leq e^{||X||} = \lim_{K\to\infty} L_K.$$

■

Proposition 3.5. *Let X be a $n \times n$ complex matrix. Then e^{tX} is a smooth curve in $GL_n(\mathbb{C})$ and*

$$\frac{d}{dt} e^{tX} = Xe^{tX} = e^{tX}X.$$

In particular,

$$\left.\frac{d}{dt}\right|_{t=0} e^{tX} = X.$$

Proof. For each i and j, we know $(e^{tX})_{ij}$ is given by an everywhere convergent power series and so we can find $\frac{d}{dt}e^{tX}$ by differentiating the power series for e^{tX} term by term. Everything else follows immediately. ■

Proposition 3.6. *Let X and Y be $n \times n$ complex matrices. Then*

$$e^{X+Y} = \lim_{m\to\infty} \left(e^{\frac{X}{m}} e^{\frac{Y}{m}} \right)^m.$$

Though this result is important, we will not prove it here, as it relies on the matrix logarithm, which we have avoided discussing due to space constraints. A good proof can be found in [1].

Definition 3.7. A function $A : \mathbb{R} \to GL_n(\mathbb{C})$ is a *one-parameter subgroup* of $GL_n(\mathbb{C})$ if

1. A is continuous,
2. $A(0) = I$,
3. $A(t + s) = A(t)A(s)$ for all $t, s \in \mathbb{R}$.

Theorem 3.8. *If A is a one-parameter subgroup of $GL_n(\mathbb{C})$, then there exists a unique $n \times n$ complex matrix X so that $A(t) = e^{tX}$.*

Though this is an important result that we will use later, we will not prove it; the proof builds upon the concept of the matrix logarithm. Skeptics should consult [1].

4 Lie Algebras

As explained in the previous section, it will be convenient to explore a Lie group's Lie algebra—its tangent space at the identity element. Such inquiry will be quite rewarding, as it will let us discover important and otherwise difficult-to-access information with ease.

Definition 4.1. A finite-dimensional real or complex *Lie algebra* is a finite-dimensional real or complex vector space $\mathfrak{g}$ together with a map $[\cdot, \cdot]$ from $\mathfrak{g} \times \mathfrak{g}$ into $\mathfrak{g}$ with the following properties:

1. $[\cdot, \cdot]$ is bilinear,
2. $[X, Y] = -[Y, X]$ for all $X, Y \in \mathfrak{g}$,
3. $[X, [Y, Z]] + [Y, [X, Z]] + [Z, [X, Y]] = 0$ for all $X, Y, Z \in \mathfrak{g}$. This property is called the *Jacobi identity*.

Definition 4.2. Let G be a matrix Lie group. The Lie algebra of G, denoted $\mathfrak{g}$, is the set of all matrices X such that $e^{tX} \in G$ for all real numbers t, and we refer to $\mathfrak{g}$ as a *matrix Lie algebra*.

These definitions may not seem to coincide; after proving the next few propositions, it will become clear that matrix Lie algebras satisfy Definition 4.1.

Proposition 4.3. *Let G be a matrix Lie group, and X an element of its Lie algebra. Then e^X is an element in the identity component of G.*

Proof. By the definition of the Lie algebra for matrix Lie groups, we know $e^{tX} \in G$ for all real numbers t. We know $A(t) = e^{tX}$ is a continuous function going from I to e^X as t goes from 0 to 1, and since I is in the identity component of G, we know e^X is as well. ■

Proposition 4.4. *Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$. Let X be an element of $\mathfrak{g}$, and A be an element of G. Then AXA^{-1} is in $\mathfrak{g}$.*

Proof. It follows from Proposition 3.4 that $e^{tAXA^{-1}} = Ae^{tX}A^{-1}$. By the definition of a Lie algebra for a matrix group we know that e^{tX} is in G for all real t, and since G is closed under inversion, we know A^{-1} is in G; thus $Ae^{tX}A^{-1} = e^{t(AXA^{-1})} \in G$, which implies $AXA^{-1} \in \mathfrak{g}$ by the definition of the Lie algebra for a matrix Lie group. ■

Now we are well-positioned to prove that matrix Lie algebras are indeed Lie algebras: as this next theorem tells us, they are vector spaces which we can equip with a bilinear antisymmetric operation satisfying the Jacobi identity.

Theorem 4.5. *Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$, and let X and Y be elements of $\mathfrak{g}$. Then*

1. *$sX \in \mathfrak{g}$ for all real numbers s,*
2. *$X + Y \in \mathfrak{g}$,*

3. $XY - YX \in \mathfrak{g}$.

Proof. Part 1. For $X \in \mathfrak{g}$, and all real t and s, we know $e^{(ts)X} \in G$. Then since $e^{(ts)X} = e^{t(sX)}$, we know that $e^{t(sX)} \in G$ for all real t and s. Then by the definition of a Lie algebra for a matrix Lie group we know $sX \in \mathfrak{g}$ for all real s.

Part 2. If $X, Y \in \mathfrak{g}$ commute, then for all real t, we know $e^{t(X+Y)} = e^{tX}e^{tY}$. Clearly e^{tX} and e^{tY} are in G for all real t, so we know $e^{t(X+Y)}$ is in G as well, which means $X + Y \in \mathfrak{g}$. In the general case, however, we use Proposition 3.6, the Lie product formula:

$$e^{t(X+Y)} = \lim_{m\to\infty} \left(e^{tX/m}e^{tY/m}\right)^m.$$

Because $X, Y \in \mathfrak{g}$, we know that $e^{tX/m}$ and $e^{tY/m}$ are in G and so is $(e^{tX/m}e^{tY/m})^m$, since G is a group. However, since G is a matrix Lie group, the limit of elements in G are also in G, so long as the limit is invertible. Since $e^{t(X+Y)}$ is invertible, we know $\lim_{m\to\infty}(e^{X/m}e^{Y/m})^m$ is in G as well. Then $\lim_{m\to\infty}(e^{tX/m}e^{tY/m})^m$ is in G, and hence so is $e^{t(X+Y)}$. This implies that $X + Y \in \mathfrak{g}$.

Part 3. Using Proposition 3.5, we can see that

$$\frac{d}{dt}e^{tX}Y\bigg|_{t=0} = XY.$$

Then by the product rule,

$$\frac{d}{dt}e^{tX}Ye^{-tX}\bigg|_{t=0} = (XY)e^0 + (e^0Y)(-X) = XY - YX.$$

By Parts 1 and 2, we know that $\mathfrak{g}$ is a vector space. By Proposition 4.4, we know $e^{tY}Xe^{-tY} \in \mathfrak{g}$ for all real t. Finally, since $\gamma(t) = e^{tX}Ye^{-tX}$ is a smooth curve through $\mathfrak{g}$, we know the derivative of γ with respect to t exists and is always in $\mathfrak{g}$; indeed, the derivative of a curve in a vector space is always in that vector space. Therefore,

$$\frac{d}{dt}\gamma(t)\bigg|_{t=0} = XY - YX$$

is in $\mathfrak{g}$. ∎

Definition 4.6. Given two $n \times n$ matrices A and B, the *commutator* of A and B, denoted $[A, B]$, is defined to be $AB - BA$.

It is easy to verify that the commutator satisfies all the necessary properties that the bracket of a Lie algebra must have. Furthermore, since the commutator of two matrices in a matrix Lie algebra is also that matrix Lie algebra, we shall use the commutator for our bracket operation when dealing with matrix Lie algebras.

It should be noted that $[\cdot, \cdot]$ is used to denote the Lie bracket of *any* Lie algebra and need not correspond to the commutator. However, since we tend to use the commutator as our Lie bracket for matrix Lie algebras, it inherits the somewhat ambiguous bracket notation.

Definition 4.7. A *subalgebra* of a real or complex Lie algebra $\mathfrak{g}$ is a subspace $\mathfrak{h}$ of $\mathfrak{g}$ such that $[H_1, H_2] \in \mathfrak{h}$ for all $H_1, H_2 \in \mathfrak{h}$. If $\mathfrak{g}$ is a complex Lie algebra and $\mathfrak{h}$ is a real subspace of $\mathfrak{g}$ closed under brackets then we say that $\mathfrak{h}$ is a real subalgebra of $\mathfrak{g}$.

If $\mathfrak{g}$ and $\mathfrak{h}$ are Lie algebras then a linear map $\phi : \mathfrak{g} \to \mathfrak{h}$ is called a *Lie algebra homomorphism* if $\phi([X, Y]) = [\phi(X), \phi(Y)]$ for all $X, Y \in \mathfrak{g}$. Furthermore, if ϕ is also bijective, then ϕ is called a *Lie algebra isomorphism*. Lastly, a Lie algebra isomorphism with Lie algebra $\mathfrak{g}$ as both its domain and codomain is called a *Lie algebra automorphism*.

Theorem 4.8. *Let G and H be Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, respectively. Suppose $\Phi : G \to H$ is a Lie group homomorphism. Then there exists a unique linear map $\phi : \mathfrak{g} \to \mathfrak{h}$ such that $\Phi(e^X) = e^{\phi(X)}$, and*

1. *$\phi(AXA^{-1}) = \Phi(A)\phi(X)\Phi(A)^{-1}$, for all $X \in \mathfrak{g}$, $A \in G$,*
2. *$\phi([X, Y]) = [\phi(X), \phi(Y)]$, for all $X, Y \in \mathfrak{g}$,*
3. *$\phi(X) = \frac{d}{dt}\Phi(e^{tX})|_{t=0}$, for all $X \in \mathfrak{g}$.*

Proof. Since Φ is a continuous group homomorphism and e^{tX} is also continuous, we know $\Phi(e^{tX})$ will be a one-parameter subgroup of H. By Theorem 3.8, we know there is a unique matrix Z so that $\Phi(e^{tX}) = e^{tZ}$ for all $t \in \mathbb{R}$. Furthermore, we know $Z \in \mathfrak{h}$, since $e^{tZ} = \Phi(e^{tX}) \in H$ for all t. Now we simply define $\phi(X) = Z$ and check that the necessary properties are satisfied.

Step 1: $\Phi(e^X) = e^{\phi(X)}$.

This follows from a few simple facts: we know $e^{tZ} = \Phi(e^{tX})$ for all t, and $\phi(X) = Z$. Thus, $e^{t\phi(X)} = \Phi(e^{tX})$ for all t, and in particular for $t = 1$, we know $e^{\phi(X)} = \Phi(e^X)$. Now for linearity!

Step 2: $\phi(sX) = s\phi(X)$.

For all $s, t \in \mathbb{R}$, we have $e^{t\phi(sX)} = \Phi\big(e^{t(sX)}\big)$, and $e^{t(s\phi(X))} = \Phi\big(e^{t(sX)}\big)$. But for the first equation, we know that $\phi(sX)$ is the unique matrix so that $e^{t\phi(sX)} = \Phi(e^{tsX})$. By the second equation, we know $s\phi(X)$ is the unique matrix so that $e^{s\phi(X)} = \Phi(e^{tsX})$. Hence, $s\phi(X) = \phi(sX)$.

Step 3: $\phi(X + Y) = \phi(X) + \phi(Y)$.

By Steps 1 and 2, we know that

$$e^{t\phi(X+Y)} = e^{\phi(t(X+Y))} = \Phi\big(e^{t(X+Y)}\big).$$

By the Lie product formula from Proposition 3.6, and the fact that Φ is a continuous homomorphism, we have

$$e^{t\phi(X+Y)} = \Phi\left(\lim_{m\to\infty}\left(e^{tX/m}e^{tY/m}\right)^m\right) = \lim_{m\to\infty}\left(\Phi(e^{tX/m})\Phi(e^{tY/m})\right)^m.$$

But then by the relationship between Φ and ϕ, and applying the Lie product formula from Proposition 3.6, we know that

$$\lim_{m\to\infty}\left(\Phi(e^{tX/m})\Phi(e^{tY/m})\right)^m = \lim_{m\to\infty}\left(e^{t\phi(X)/m}e^{t\phi(Y)/m}\right)^m = e^{t(\phi(X)+\phi(Y))}.$$

Thus, $e^{t\phi(X+Y)} = e^{t(\phi(X)+\phi(Y))}$. Using Proposition 3.5, we can differentiate both of these at $t = 0$ to get $\phi(X + Y) = \phi(X) + \phi(Y)$.

Step 4: $\phi(AXA^{-1}) = \Phi(A)\phi(X)\Phi(A)^{-1}$.

By Steps 1 and 2,

$$e^{t\phi(AXA^{-1})} = e^{\phi(tAXA^{-1})} = \Phi(e^{tAXA^{-1}}).$$

By Proposition 3.4 and Step 1, we know that

$$e^{t\phi(AXA^{-1})} = \Phi(e^{tAXA^{-1}}) = \Phi(Ae^{tX}A^{-1}) = \Phi(A)\Phi(e^{tX})\Phi(A^{-1}).$$

And since we know that $\Phi(A^{-1}) = \Phi(A)^{-1}$ for any homomorphism Φ, and since $\Phi(e^{tX}) = e^{t\phi(X)}$, we know

$$e^{t\phi(AXA^{-1})} = \Phi(A)e^{t\phi(X)}\Phi(A)^{-1}.$$

Differentiating at $t = 0$ we obtain

$$\phi(AXA^{-1}) = \Phi(A)\phi(X)\Phi(A)^{-1}.$$

Step 5: $\phi([X, Y]) = [\phi(X), \phi(Y)]$.

Recall from the proof of Theorem 4.5, we know that

$$[X, Y] = \frac{d}{dt}e^{tX}Ye^{-tX}\bigg|_{t=0}.$$

Hence,

$$\phi([X, Y]) = \phi\left(\frac{d}{dt}e^{tX}Ye^{-tX}\bigg|_{t=0}\right) = \frac{d}{dt}\phi(e^{tX}Ye^{-tX})\bigg|_{t=0},$$

since a derivative commutes with a linear transformation. Then by Step 1,

$$\phi([X, Y]) = \frac{d}{dt}\Phi(e^{tX})\phi(Y)\Phi(e^{-tX})\bigg|_{t=0} = \frac{d}{dt}e^{t\phi(X)}\phi(Y)e^{-t\phi(X)}\bigg|_{t=0}.$$

Of course, we know from the proof of Theorem 4.5 that the far right side of this equation is equal to $\phi(X)\phi(Y) - \phi(Y)\phi(X)$, so

$$\phi([X, Y]) = [\phi(X), \phi(Y)],$$

and thus ϕ is a Lie algebra homomorphism.

Step 6: $\phi(X) = \frac{d}{dt}\Phi(e^{tX})\big|_{t=0}$.

To begin with, it is clear that $e^{t\phi(X)} = \Phi(e^{tX})$, so

$$\frac{d}{dt}\Phi(e^{tX})|_{t=0} = \frac{d}{dt}e^{t\phi(X)}\bigg|_{t=0}.$$

By Proposition 3.5, we know $\frac{d}{dt}e^{t\phi(X)}\Big|_{t=0} = \phi(X)$, so

$$\phi(X) = \frac{d}{dt}\Phi(e^{tX})\Big|_{t=0}.$$

Step 7: ϕ is the unique linear map such that $\Phi(e^{tX}) = e^{t\phi(X)}$.
Suppose ψ is another such linear map. Then,

$$e^{t\psi(X)} = e^{\psi(tX)} = \Phi(e^{tX}).$$

And so by Step 6,

$$\psi(X) = \frac{d}{dt}\Phi(e^{tX})\Big|_{t=0} = \phi(X).$$

■

Theorem 4.9. *Suppose that G, H, and K are matrix Lie groups, with corresponding Lie algebras $\mathfrak{g}$, $\mathfrak{h}$, and $\mathfrak{k}$. Let $\Phi : H \to K$ and $\Psi : G \to H$ be Lie group homomorphisms, and let $\Lambda : G \to K$ be the composition of Φ and Ψ, so that $\Lambda(A) = \Phi(\Psi(A))$ for all A in G. Let $\phi : \mathfrak{h} \to \mathfrak{k}$, $\psi : \mathfrak{g} \to \mathfrak{h}$, and $\lambda : \mathfrak{g} \to \mathfrak{k}$ be the associated Lie algebra homomorphisms such that $e^{\phi(X)} = \Phi(e^X)$, $e^{\psi(X)} = \Psi(e^X)$, and $e^{\lambda(X)} = \Lambda(e^X)$. Then for all $X \in \mathfrak{g}$, $\lambda(X) = \phi(\psi(X))$.*

Proof. For any $X \in \mathfrak{g}$,

$$e^{t\lambda(X)} = \Lambda(e^{tX}) = \Phi\big(\Psi(e^{tX})\big) = \Phi\big(e^{t\psi(X)}\big) = e^{t\phi(\psi(X))}.$$

Differentiating at $t = 0$, we know by Proposition 3.5 that

$$\lambda(X) = \frac{d}{dt}e^{t\lambda(X)}\Big|_{t=0} = \frac{d}{dt}e^{t\phi(\psi(X))}\Big|_{t=0} = \phi(\psi(X)).$$

■

Definition 4.10. Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$. Then for each $A \in G$ define a linear map $Ad_A : \mathfrak{g} \to \mathfrak{g}$ by the formula $Ad_A(X) = AXA^{-1}$. This map is called the *adjoint mapping*.

Proposition 4.11. *Let G be a matrix Lie group, with Lie algebra $\mathfrak{g}$. Let $GL(\mathfrak{g})$ denote the group of all invertible linear transformations of $\mathfrak{g}$. Then for each $A \in G$, Ad_A is an invertible linear transformation of $\mathfrak{g}$ with inverse $Ad_{A^{-1}}$, and the map $Ad : A \mapsto Ad_A$ is a group homomorphism of G into $GL(\mathfrak{g})$. Furthermore, for each $A \in G$, Ad_A satisfies $Ad_A([X,Y]) = [Ad_A(X), Ad_A(Y)]$ for all $X, Y \in \mathfrak{g}$.*

Proof. We can see that for any $X \in \mathfrak{g}$, and any $A \in G$,

$$Ad_A(Ad_{A^{-1}}(X)) = A(A^{-1}XA)A^{-1} = X = A^{-1}(AXA^{-1})A = Ad_{A^{-1}}(Ad_A(X)).$$

And now

$$Ad(AB) = Ad_{AB}(\,\cdot\,) = AB(\,\cdot\,)B^{-1}A^{-1} = Ad_A(Ad_B(\,\cdot\,)) = Ad(A)(Ad(B)).$$

Since multiplication of matrices is the group operation of G and composition of linear maps is the group operation in $GL(\mathfrak{g})$, we know that the Ad operator is a group homomorphism. And lastly, for any $X, Y \in \mathfrak{g}$ and any $A \in G$,

$$Ad_A([X,Y]) = A(XY - YX)A^{-1} = AXYA^{-1} - AYXA^{-1}.$$

And the far right side of this equality is clearly

$$AXA^{-1}AYA^1 - AYA^{-1}AXA^{-1} = Ad_A(X)Ad_A(Y) - Ad_A(Y)Ad_A(X).$$

Thus $Ad_A([X,Y]) = [Ad_A(X), Ad_A(Y)]$, and so the map Ad_A is a Lie algebra homomorphism for each $A \in G$. ■

Since $\mathfrak{g}$ is a real vector space with some dimension k, we can pick a basis for $\mathfrak{g}$, and $GL(\mathfrak{g})$ can be written as a group of matrices with the group operation being matrix multiplication. (This notion is consistent with the traditional meaning of the general linear group of a vector space as the group of automorphisms on that vector space, having composition as the group operation.) Thus we can regard $GL(\mathfrak{g})$ as a matrix Lie group. It is easy to show that $Ad : G \to GL(\mathfrak{g})$ is continuous and so is a Lie group homomorphism. By Theorem 4.8, there is an associated real linear map ad taking X to ad_X from the Lie algebra of G to the Lie algebra of $GL(\mathfrak{g})$ (that is, from $\mathfrak{g}$ to $\mathfrak{gl}(\mathfrak{g})$, the space of all endomorphisms of $\mathfrak{g}$). Specifically, ad satisfies

$$e^{ad_X} = Ad(e^X).$$

Proposition 4.12. *Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$. Let $Ad : G \to GL(\mathfrak{g})$ be the Lie group homomorphism defined above. Let $ad : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ be the associated Lie algebra map. Then for all $X, Y \in \mathfrak{g}$, $ad_X(Y) = [X,Y]$.*

Proof. By Theorem 4.8, we know that ad can be calculated by

$$ad_X = \frac{d}{dt} Ad(e^{tX})\bigg|_{t=0}.$$

Thus,

$$ad_X(Y) = \frac{d}{dt} Ad(e^{tX})(Y)\bigg|_{t=0} = \frac{d}{dt} e^{tX}Ye^{-tX}\bigg|_{t=0} = [X,Y].$$

■

Definition 4.13. A *representation of a Lie group G* on a vector space V is a Lie group homomorphism $\rho : G \to GL(V)$, sending elements of G to automorphisms on V. Similarly, a *representation of a Lie algebra* $\mathfrak{g}$ on a vector space V is a Lie algebra homomorphism $\rho : \mathfrak{g} \to \mathfrak{gl}(V)$, mapping elements of $\mathfrak{g}$ to endomorphisms on V.

Though we will not prove it here, it turns out that all finite-dimensional Lie algebras can be represented with matrices.

Theorem (Ado) 4.14. *Every finite-dimensional real Lie algebra is isomorphic to a real subalgebra of* $\mathfrak{gl}_n(\mathbb{R})$. *Every finite-dimensional complex Lie algebra is isomorphic to a complex subalgebra of* $\mathfrak{gl}_n(\mathbb{C})$.

5 Universal Enveloping Algebras

Generally speaking, a Lie algebra $\mathfrak{g}$ does not have any defined notion of associative multiplication. However, if we consider a representation $\rho : \mathfrak{g} \to \mathfrak{gl}(V)$ then the product $\rho(X)\rho(Y)$ is well-defined. (Note that the "product" is actually composition of the operators $\rho(X)$ and $\rho(Y)$, which are endomorphisms of V.)

With convenience of multiplication and the structure passed on by the commutator, we will define the notion of the "universal" associative algebra generated by "products" of operators of the form $\rho(X)$ for $X \in \mathfrak{g}$. To make things both more formal and more interesting, we will first introduce a few new concepts:

Definition 5.1. An *associative algebra* A is a vector space V over a field $\mathbb{K}$ equipped with an associative, bilinear vector product $\cdot : V \times V \to V$. If there is some element $1 \in V$ such that $1 \cdot a = a = a \cdot 1$ for every $a \in A$, then we say that A is unital, or "has unit." Often we will describe an associative algebra A as "having unit over $\mathbb{K}$" to mean that it is a unital algebra with its vector space over a field $\mathbb{K}$.

In particular, we will be very interested in an associative algebra called the tensor algebra. But first, we must define the tensor product.

Definition 5.2. Let V_p for $1 \leq p \leq k$ and W be modules over a ring R. A module $\mathcal{T} = \mathcal{T}_{V_1,\cdots V_k}$ together with a multilinear map $\otimes : V_1 \times \cdots \times V_k \to \mathcal{T}$ is called *universal for k-multilinear maps on* $V_1 \times \cdots \times V_k$ if for every multilinear map $\mu : V_1 \times \cdots V_k \to W$ there is a unique linear map $\tilde{\mu} : \mathcal{T} \to W$ such that $\tilde{\mu} \circ \otimes = \mu$. If such a universal object exists, it will be called a *tensor product*.

It turns out that the tensor product is unique up to isomorphism.

Proposition 5.3. *If* $(\mathcal{T}_1, \otimes_1)$ *and* $(\mathcal{T}_2, \otimes_2)$ *are both universal for k-multilinear maps on* $V_1 \times \cdots \times V_k$, *then there is a unique isomorphism* $\Phi : \mathcal{T}_1 \to \mathcal{T}_2$ *such that* $\Phi \circ \otimes_1 = \otimes_2$.

Proof. By the assumption of universality, we know that there are maps $\otimes_1$ and $\otimes_2$ such that $\Phi \circ \otimes_1 = \otimes_2$, and $\bar{\Phi} \circ \otimes_2 = \otimes_1$. Thus we have $\bar{\Phi} \circ \Phi \circ \otimes_1 = \otimes_1$, and by the uniqueness part of the universality of $\otimes_1$ it follows that $\bar{\Phi} \circ \Phi = Id$. Similarly, $\Phi \circ \bar{\Phi} \circ \otimes_2 = \otimes_2$, and by the uniqueness part of the universality of $\otimes_2$ it follows that $\Phi \circ \bar{\Phi} = Id$. Thus $\bar{\Phi} = \Phi^{-1}$. ■

More concretely, the realization of the tensor product of modules $V_1, \cdots V_k$ is, at least roughly, the set of all linear combinations of symbols of the form $v_1 \otimes \cdots \otimes v_k$ subject to the multilinear relations

$$v_1 \otimes \cdots \otimes av_i \otimes \cdots \otimes v_k = a(v_1 \otimes \cdots \otimes v_i \otimes \cdots \otimes v_k),$$

and

$$v_1 \otimes \cdots \otimes (v_i + v_i') \otimes \cdots \otimes v_k = (v_1 \otimes \cdots \otimes v_i \otimes \cdots \otimes v_k) + (v_1 \otimes \cdots \otimes v_i' \otimes \cdots \otimes v_k).$$

This space is denoted by $V_1 \otimes \cdots \otimes V_k$, and is called the tensor product of the modules $V_1, \cdots, V_k$.

Definition 5.4. The *tensor algebra* of a vector space V over a field $\mathbb{K}$, denoted $T(V)$, is the associative algebra of tensors on V, with the tensor product $\otimes$ serving as the associative, bilinear vector product.

For any vector space, we can construct its tensor algebra as follows:

Let $T^0V = \mathbb{K}$. For any $k \in \mathbb{N}$, define the k-th tensor power of V, denoted T^kV, to be the tensor product of V with itself k times:

$$T^kV = V \otimes V \otimes \cdots \otimes V.$$

From here, we simply take the direct sum of the T^kV for $k = 0, 1, 2, \ldots,$

$$T(V) = \bigoplus_{k=0}^{\infty} T^kV = \mathbb{K} \oplus V \oplus (V \otimes V) \oplus (V \otimes V \otimes V) \cdots .$$

Now we are in a suitable position to discuss what is known as the "universal property" of tensor algebras.

Proposition 5.5. *Let A be any associative algebra with unit over $\mathbb{K}$, and let $f : V \to A$ be a linear map. Then there exists a unique algebra homomorphism $\bar{f} : T(V) \to A$ such that $f = \bar{f} \circ i$, where $i : V \to T(V)$ is the inclusion of $V = T^1V$ into $T(V)$.*

Proof. Let A be any associative algebra with unit over $\mathbb{K}$. For any linear map $f : V \to A$, define a linear map $\bar{f} : T(V) \to A$ by

$$\bar{f}(v_1 \otimes \cdots \otimes v_k) = f(v_1) \cdots f(v_k).$$

This will be well-defined for any $k \in \mathbb{N}$, and it is easy to see that it is indeed an algebra homomorphism. However, if $k = 0$, then we must clarify. Fortunately, this is easy: since A is an algebra with unit over $\mathbb{K}$, we can simply let $\bar{f}(\ell) = \ell \cdot 1$, where 1 is the unit element of A and $\ell \in \mathbb{K}$. The above definition is clearly the only way to extend f as a homomorphism, since V generates $T(V)$ as a $\mathbb{K}$-algebra. ∎

Now for one more detail that will be important later:

Definition 5.6. Let I be the two-sided ideal in $T(V)$ generated by all the $X \otimes Y - Y \otimes X$ (where $X, Y \in V$). Define the *symmetric algebra* of V, denoted $\mathrm{Sym}(V)$, by $\mathrm{Sym}(V) = T(V)/I$.

In other words, the symmetric algebra of V is just the commutative version of the tensor algebra. With that out of the way, we are ready to move on to the real topic of this section.

Definition 5.7. Let $\mathfrak{g}$ be a Lie algebra over a field $\mathbb{K}$. The *universal enveloping algebra* of $\mathfrak{g}$ is a pair $(\mathfrak{Ug}, i)$, satisfying the following:

1. $\mathfrak{Ug}$ is an associative algebra with unit over $\mathbb{K}$,
2. $i : \mathfrak{g} \to \mathfrak{Ug}$ is linear and $i(X)i(Y) - i(Y)i(X) = i([X, Y])$, for all $X, Y \in \mathfrak{g}$,
3. for any associative algebra A with unit over $\mathbb{K}$ and for any linear map $j : \mathfrak{g} \to A$ satisfying $j(X)j(Y) - j(Y)j(X) = j([X, Y])$ for each $X, Y \in \mathfrak{g}$, there exists a unique homomorphism of algebras $\phi : \mathfrak{Ug} \to A$ such that $\phi \circ i = j$.

Notice that the Lie bracket is not necessarily the commutator—after all, there may be no notion of associative multiplication in $\mathfrak{g}$—but that applying i to the bracket of any two $X, Y \in \mathfrak{g}$ must give us the commutator of $i(X)$ and $i(Y)$.

Theorem 5.8. *For any Lie algebra $\mathfrak{g}$ over an arbitrary field $\mathbb{K}$, there exists a unique universal enveloping algebra $(\mathfrak{Ug}, i)$, up to isomorphism.*

Proof. Uniqueness: Suppose that the Lie algebra $\mathfrak{g}$ has two universal enveloping algebras $(\mathfrak{Ug}, i)$ and $(\mathfrak{Ug}', i')$. Then by definition, for each associative $\mathbb{K}$-algebra A there exists a unique homomorphism $\lambda_A : \mathfrak{Ug} \to A$. In particular, since $\mathfrak{Ug}'$ is an associative $\mathbb{K}$-algebra, we have a unique homomorphism of algebras $\lambda : \mathfrak{Ug} \to \mathfrak{Ug}'$. Switching the roles of $\mathfrak{Ug}$ and $\mathfrak{Ug}'$ and applying the same logic, we know there exists a unique homomorphism of algebras $\mu : \mathfrak{Ug}' \to \mathfrak{Ug}$. Then $\lambda \circ \mu = 1_{\mathfrak{Ug}}$, and $\mu \circ \lambda = 1_{\mathfrak{Ug}'}$, which means λ is bijective. But a bijective homomorphism of algebras is an isomorphism of algebras. Thus $(\mathfrak{Ug}, i)$ is unique up to isomorphism.

Existence: Let $T(\mathfrak{g})$ be the tensor algebra of $\mathfrak{g}$, and let J be the two-sided ideal in $T(\mathfrak{g})$ generated by all $X \otimes Y - Y \otimes X - [X, Y]$, where $X, Y \in \mathfrak{g}$. We claim that $\mathfrak{Ug} = T(\mathfrak{g})/J$ satisfies all the necessary conditions delineated in Definition 5.7. Let $\pi : T(\mathfrak{g}) \to \mathfrak{Ug}$ be the homomorphism mapping each element of the tensor algebra to its equivalence class in the associative algebra $T(\mathfrak{g})/J$. Clearly,

$$J \subset \bigoplus_{k>0} T^k\mathfrak{g}.$$

It follows that π maps $T^0\mathfrak{g} = \mathbb{K}$ isomorphically into $T(\mathfrak{g})/J$, and hence $\mathfrak{Ug}$ at least contains scalars—great! Now let $i : \mathfrak{g} \to \mathfrak{Ug}$ be the restriction of π to $\mathfrak{g} \subset T(\mathfrak{g})$. Let A be any associative algebra with unit over $\mathbb{K}$, and let $j : \mathfrak{g} \to A$ be a linear map satisfying $j(X)j(Y) - j(Y)j(X) = j([X, Y])$ for all $X, Y \in \mathfrak{g}$.

The universal property of tensor algebras from Proposition 5.5 gives us a unique algebra homomorphism $\phi' : T(\mathfrak{g}) \to A$ that extends j and sends 1 to 1. It follows from the special property of j that $X \otimes Y - Y \otimes X - [X, Y]$ is in $\mathrm{Ker}(\phi')$ for all $X, Y \in \mathfrak{g}$. Thus, since each of the elements in $T(\mathfrak{g})$ generated by the terms $X \otimes Y - Y \otimes X - [X, Y]$ gets mapped to zero, we can identify all such elements and a homomorphism shall still exist. In other words, ϕ' induces a homomorphism $\phi : \mathfrak{Ug} \to A$ such that $\phi \circ i = j$. The uniqueness of ϕ is evident, since 1 and $\mathrm{Im}(i)$ together generate $\mathfrak{Ug}$. ■

Now for a simple example. If $\mathfrak{g}$ is an abelian Lie algebra (i.e., $[X, Y] = 0$ for all $X, Y \in \mathfrak{g}$), then $\mathfrak{Ug} = T(\mathfrak{g})/J$, where J is the two-sided ideal generated by all $X \otimes Y - Y \otimes X - [X, Y]$. But since $[X, Y]$ is always zero, we simply have $\mathfrak{Ug} = \mathrm{Sym}(\mathfrak{g})$.

6 The Poincaré-Birkhoff-Witt Theorem

It turns out that $\mathfrak{g}$ is mapped injectively into $\mathfrak{Ug}$, and hence universal enveloping algebras are indeed "enveloping" in some sense. This becomes quite useful, since we can proceed think of each $i(X)$ as simply being the corresponding $X \in \mathfrak{g}$; the lack of restrictions that characterize the universal enveloping algebra allow us to perform feats that would be cumbersome or impossible in the Lie algebra itself. This important result turns out to be a simple corollary of the much stronger Poincaré-Birkhoff-Witt Theorem.

Definition 6.1. A *graded algebra* $\mathfrak{G}$ is an associative algebra that can be decomposed into the direct sum of abelian groups G_k (with the group operation being addition of vectors) and is characterized by the fact that if $X \in G_m$ and $Y \in G_p$ then $X \cdot Y \in G_{m+p}$.

For instance, the tensor algebra $T(V)$ of any vector space V is a graded algebra. Clearly if $X \in T^m V$ and $Y \in T^p V$, then $X \otimes Y \in T^{m+p} V$. If we mod out by the two-sided ideal I generated by the $X \otimes Y - Y \otimes X$, for $X, Y \in V$, we get $\mathrm{Sym}(V) = T(V)/I$. This is a graded algebra as well, as we can define a grading $S^m V = T^m V / I$.

Definition 6.2. A *filtration* is an indexed set Q_i of subobjects in a given algebraic structure Q, where the index i runs over an index set S which is totally ordered, and if $i \leq j$ then $Q_i \subset Q_j$.

Definition 6.3. A *filtered algebra* $\mathfrak{F}$ is an associative algebra over some field $\mathbb{K}$, which has a filtration of linear subspaces, indexed by a set S, satisfying $\{0\} \subset F_0 \subset F_1 \subset \cdots \subset \mathfrak{F}$, recovering $\mathfrak{F}$ via the union

$$\bigcup_{s \in S} F_s = \mathfrak{F},$$

and satisfying the following: if $X \in F_m$ and $Y \in F_p$, then $X \cdot Y \in F_{m+p}$.

So far, we know very little about $\mathfrak{Ug}$, other than the fact that it contains the scalars which were passed on isomorphically from $\mathbb{K}$. For brevity, we shall write T instead if $T(\mathfrak{g})$, and Sym instead of $\mathrm{Sym}(\mathfrak{g})$. Similarly, we shall write T^m instead of $T^m\mathfrak{g}$, and S^m instead of $S^m\mathfrak{g}$. For algebras other than the tensor algebra itself, we shall also frequently omit the $\otimes$, choosing a dot or simply placing variables next to each other to indicate multiplication.

Define a filtration on T by

$$T_m := T^0 \oplus T^1 \oplus \cdots \oplus T^m.$$

Recall that $\pi : T \to \mathfrak{Ug}$ is the quotient map. Let $U_m = \pi(T_m)$, and $U_{-1} = 0$. Suppose we have $W \in T_m$, $Z \in T_p$, and define $X = \pi(W) \in U_m$, and $Y = \pi(Z) \in U_p$. Then $W \otimes Z \in T_{m+p}$, which implies that $\pi(W \otimes Z) \in \pi(T_{m+p})$, and hence

$$XY = \pi(W)\pi(Z) = \pi(W \otimes Z) \in \pi(T_{m+p}) = U_{m+p}.$$

Thus, for all $X \in U_m, Y \in U_p$, $XY \in U_{m+p}$, so the U_m's form a filtration on $\mathfrak{Ug}$.

Define

$$G^m := U_m/U_{m-1}$$

(this is just a vector space), and let the multiplication in $\mathfrak{Ug}$ define a bilinear map $G^m \times G^p \to G^{m+p}$. This operation is well defined, as we shall see momentarily. Suppose we have two representatives $X, X' \in U_p$ of the same equivalence class in G^p, and two representatives $Y, Y' \in U_m$ of the same equivalence class in G^m. Define $W = X - X' \in U_{p-1}$ and $Z = Y - Y' \in U_{m-1}$. Then

$$XY = (X' + W)(Y' + Z) = X'Y' + (X'Z + WY' + WZ).$$

But surely $WY' \in U_{m+p-1}$, $X'Z \in U_{m+p-1}$, and $WZ \in U_{(m-1)+(p-1)} \subset U_{m+p-1}$. Thus, when we mod out by U_{m+p-1}, in accordance with our definition of G^{m+p}, all the terms in the parentheses vanish. Having gotten that out of the way, we define

$$\mathfrak{G} = \bigoplus_{m=0}^{\infty} G^m.$$

This gives us a bilinear map $\mathfrak{G} \times \mathfrak{G} \to \mathfrak{G}$ in accordance with the rules of multiplication for $G^m \times G^p \to G^{m+p}$. It is clear that $\mathfrak{G}$ is a graded associative algebra with unit.

Since π maps the last bit T^m of each T_m into U_m, it follows that the composite linear map $\phi_m : T^m \to U_m \to G^m = U_m/U_{m-1}$ is well defined. And we can write $T_m = T_{m-1} \oplus T^m$. Clearly, π maps T_m surjectively onto $U_m = \pi(T_m)$. And surely the mapping from U_m to U_m/U_{m-1} is surjective, and so the map $\Phi_m : T_m \to U_m \to U_m/U_{m-1}$ is surjective as well. And since this map maps everything in T_{m-1} to $U_{m-1}/U_{m-1} = \{0\}$, we know that the restriction of Φ_m to T^m, which we call $\phi_m : T^m \to U_m \to U_m/U_{m-1}$, hits everything in U_m/U_{m-1}, except possibly zero. But we know $0 = 0 \otimes \cdots \otimes 0 \in T^m$, so zero is in the image of T^m under ϕ_m. Hence, ϕ_m is surjective onto G^m. The maps ϕ_m therefore can be combined to give us a surjective linear map $\phi : T \to \mathfrak{G}$ sending 1 to 1.

Lemma 6.4. *The map $\phi : T \to \mathfrak{G}$ is an algebra homomorphism. Moreover, if I is the two-sided ideal generated by $X \otimes Y - Y \otimes X$ for $X, Y \in \mathfrak{g}$, then $I \in \mathrm{Ker}(\phi)$, and so ϕ induces a homomorphism ω of $\mathrm{Sym} = T/I$ onto $\mathfrak{G}$.*

Proof. Suppose we have some $X \in T^m$ and $Y \in T^p$. It follows that $\phi(X) \in G^m$, $\phi(Y) \in G^p$, and $X \otimes Y \in T^{m+p}$, so $\phi(X \otimes Y) \in G^{m+p}$. Then by the definition of the product in $\mathfrak{G}$, it follows that $\phi(X \otimes Y) = \phi(X)\phi(Y)$ for each $X, Y \in T$. Thus, ϕ is a (surjective) algebra homomorphism.

Let $X \otimes Y - Y \otimes X$ (for $X, Y \in \mathfrak{g}$) be a typical generator of the two-sided ideal I described earlier. Then $\pi(X \otimes Y - Y \otimes X) \in U_2$, by definition. On the other hand, we also know $\pi(X \otimes Y - Y \otimes X) = \pi([X, Y]) \in U_1$, and so $\phi(X \otimes Y - Y \otimes X) \in U_1/U_1 = \{0\}$. Hence $I \subset \mathrm{Ker}(\phi)$, and so if we identify all the elements of the ideal I with the zero vector, we surely shall still have a (surjective) algebra homomorphism $\omega : \mathrm{Sym} \to \mathfrak{G}$. ∎

It turns out that ω is not only a surjective algebra homomorphism, but is also injective, and hence is an isomorphism of algebras. This fundamental result is known as the Poincaré-Birkhoff-Witt Theorem, which we shall prove after introducing the important corollaries it entails.

Theorem (Poincaré-Birkhoff-Witt) 6.5. *The homomorphism $\omega : \mathrm{Sym} \to \mathfrak{G}$ is an isomorphism of algebras.*

Corollary 6.6. *Let W be a subspace of T^m. Suppose the canonical map $T^m \to S^m$ sends W isomorphically onto S^m. Then $\pi(W)$ is a complement to U_{m-1} in U_m.*

Proof. Let g_m be the quotient map from T^m to S^m, and let h_m be the quotient map from U_m to U_m/U_{m-1}. By Lemma 6.4, and the definitions, we know the diagram below is commutative:

$$\begin{array}{ccc} T^m & \xrightarrow{\pi} & U_m \\ {\scriptstyle g_m}\downarrow & & \downarrow{\scriptstyle h_m} \\ S^m & \xrightarrow[\omega]{} & G^m \end{array}$$

Since g_m sends $W \subset T^m$ isomorphically onto S^m by our supposition, and since $\omega : \mathrm{Sym} \to \mathfrak{G}$ is an isomorphism by the Poincaré-Birkoff-Witt Theorem, we know the map $\omega \circ g_m$ sends W isomorphically onto G^m. Since W is mapped isomorphically, it is mapped injectively, and hence $\mathrm{Ker}(h_m \circ \pi) \cap W = \{0\}$. It follows that $\mathrm{Ker}(h_m) \cap \pi(W) = \{0\}$ as well. But the kernel of h_m is just U_{m-1}, and so $U_{m-1} \cap \pi(W) = \{0\}$.

And since W is mapped isomorphically onto G^m, we know that h_m is an isomorphism from $\pi(W)$ to $h_m(\pi(W)) = U_m/U_{m-1} = h_m(U_m)$. By the Rank-Nullity Theorem,

$$U_m \cong \mathrm{Ker}(h_m) \oplus \mathrm{Im}(h_m).$$

The kernel of h_m is just U_{m-1}, and in this context $\mathrm{Im}(h_m) = h_m(U_m) = h_m(\pi(W)) \cong \pi(W)$. Hence

$$U_m \cong U_{m-1} \oplus \pi(W). \qquad \blacksquare$$

Corollary 6.7. *The canonical map $i : \mathfrak{g} \to \mathfrak{Ug}$ is injective.*

Proof. This is just a special case of Corollary 6.6, with $m = 1$, and $W = T^1 = \mathfrak{g}$. The supposition in Corollary 6.6 tells us that T^1 and S^1 are isomorphic holds because we have $S^1 = g_m(T^1) = T^1$. Using the same logic as in the proof of Corollary 6.6, it is clear that $W = \mathfrak{g}$ must be mapped isomorphically onto U_1. Since U_1 is in the filtration of $\mathfrak{Ug}$ we know that $\mathfrak{g}$ is mapped injectively into $\mathfrak{Ug}$ itself. $\blacksquare$

This result is very important—it allows us to identity each $X \in \mathfrak{g}$ with $i(X) \in \mathfrak{Ug}$, and hence think of $\mathfrak{Ug}$ as a bigger algebra "enveloping" the Lie algebra $\mathfrak{g}$.

Corollary 6.8. *Let $(x_1, x_2, x_3, \cdots)$ be any ordered basis of $\mathfrak{g}$. Then the elements $x_{i_1} \cdots x_{i_m} = \pi(x_{i_1} \otimes \cdots \otimes x_{i_m})$, where $m \in \mathbb{N}$, and $i_1 \leq i_2 \leq \cdots \leq i_m$, along with 1, form a basis of $\mathfrak{Ug}$.*

Proof. Let W be the subspace of T^m spanned by all $x_{i_1} \otimes \cdots \otimes x_{i_m}$, where $i_1 \leq i_2 \leq \cdots \leq i_m$. It is evident that W maps isomorphically onto S^m, so we know by Corollary 6.6 that $\pi(W)$ is a complement to U_{m-1} in U_m, and it is easy to see by induction that the union of 1 and the bases of each U_m, for $m \in \mathbb{N}$, form a basis for $\mathfrak{Ug}$. $\blacksquare$

And now we set out to prove the Poincaré-Birkhoff-Witt Theorem itself. But before we do, we shall introduce some new notation.

Fix an ordered basis $(x_\lambda : \lambda \in \Omega)$ of $\mathfrak{g}$. This choice identifies Sym with the polynomial algebra in indeterminates z_λ, where $\lambda \in \Omega$. For each sequence $\Sigma = (\lambda_1, \lambda_2, \cdots \lambda_m)$ of indices (m is called the length of Σ), let $z_\Sigma = z_{\lambda_1} \cdots z_{\lambda_m} \in S^m$ and let $x_\Sigma = x_{\lambda_1} \otimes \cdots \otimes x_{\lambda_m} \in T^m$. We shall call Σ increasing if $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m$ in a given ordering of Ω. By fiat, let $\varnothing$ be increasing and set $z_\varnothing = 1$. It follows that the set $\{z_\Sigma \,|\, \Sigma \text{ increasing}\}$ is a basis of Sym. Later on, the fact that Sym is a filtered algebra will be of importance, and so we note it now: associated with the grading of

$$\mathrm{Sym} = \bigoplus_{k=0}^{\infty} S^k$$

is a filtration

$$S_k = S^0 \oplus \cdots \oplus S^k.$$

Lastly, in the following lemmas, we shall write $\lambda \leq \Sigma$ if $\lambda \leq \mu$ for all $\mu \in \Sigma$.

Lemma 6.9. *For each $m \in \mathbb{Z}^+$, there exists a unique linear map $f_m : \mathfrak{g} \otimes S_m \to$* Sym *satisfying:*

(A_m) $f_m(x_\lambda \otimes z_\Sigma) = z_\lambda z_\Sigma$ *for any $\lambda \leq \Sigma$, and any $z_\Sigma \in S_m$,*

(B_m) $f_m(x_\lambda \otimes z_\Sigma) - z_\lambda z_\Sigma \in S_k$ *for any $k \leq m$, and any $z_\Sigma \in S_k$,*

(C_m) $f_m(x_\lambda \otimes f_m(x_\mu \otimes z_T)) = f_m(x_\mu \otimes f_m(x_\lambda \otimes z_T)) + f_m([x_\lambda, x_\mu] \otimes z_T)$ *for all $z_T \in S_{m-1}$.*

Moreover, the restriction of f_m to $g \otimes S_{m-1}$ agrees with f_{m-1}.

Proof. First, note that all of the terms in (C_m) make sense once we have proven (B_m). Note further that the restriction of f_m to $\mathfrak{g} \otimes S_{m-1}$ must satisfy (A_{m-1}), (B_{m-1}), and (C_{m-1}), so this restricted map must be the same as f_{m-1} due to the asserted uniqueness. To verify that existence and uniqueness hold for each f_m, we proceed by induction on m. For $m = 0$, only $z_\Sigma = z_\varnothing = 1$ occurs; thus, we can let $f_0(x_\lambda \otimes 1) = z_\lambda$ and extend linearly to $\mathfrak{g} \otimes S_0$. Evidently, (A_0), (B_0), and (C_0) are satisfied. Furthermore, (A_0) shows that our choice of f_0 is the only possible one.

Assuming the existence of a unique f_{m-1} satisfying (A_{m-1}), (B_{m-1}), and (C_{m-1}), we will show how to extend f_{m-1} to a map f_m. For this purpose, it will suffice to define $f_m(x_\lambda \otimes z_\Sigma)$ where Σ is an increasing sequence of length m.

For the case where $\lambda \leq \Sigma$, condition (A_m) cannot hold unless we define $f_m(x_\lambda \otimes z_\Sigma) = z_\lambda z_\Sigma$. On the other hand, if $\lambda \leq \Sigma$ fails to hold, then λ is greater than some element of Σ. Certainly, then, the first index μ in Σ is strictly less than λ, and $\Sigma = (\mu, T)$, where $\mu \leq T$ and T is of length $m - 1$. By (A_{m-1}), we know $z_\Sigma = z_\mu z_T = f_{m-1}(x_\mu \otimes z_T)$. Since $\mu \leq T$, $f_m(x_\mu \otimes z_T) = z_\mu z_T = z_\Sigma$ is already defined, so the left side of (C_m) becomes $f_m(x_\lambda \otimes z_\Sigma)$. On the other hand, (B_{m-1}), with $k = m - 1$ implies that

$$f_m(x_\lambda \otimes z_T) = f_{m-1}(x_\lambda \otimes z_T) = z_\lambda z_T + y$$

for a specific $y \in S_{m-1}$, since f_{m-1} is defined uniquely by the induction hypothesis. This shows that the right side of (C_m) is already defined:

$$z_\mu z_\lambda z_T + f_{m-1}(x_\mu \otimes y) + f_{m-1}([x_\lambda, x_\mu] \otimes z_T),$$

where $y \in S_{m-1}$.

The preceding remarks show that f_m can be defined, and in only one way. Moreover, (A_m) and (B_m) certainly hold, as does (C_m), as long as $\mu < \lambda$, $\mu \leq T$. But $[x_\mu, x_\lambda] = -[x_\lambda, x_\mu]$, so (C_m) also holds for $\lambda < \mu$, $\lambda \leq T$. When $\lambda = \mu$, (C_m) also holds. We now only need to consider the case where neither $\lambda \leq T$ nor $\mu \leq T$ is true. Write $T = (v, \Psi)$, where $v \leq \Psi$, $v < \lambda$, and $v < \mu$. To keep notation under control, write $f_m(x \otimes z)$ as xz for any $x \in \mathfrak{g}$ and $z \in S_m$.

The induction hypothesis guarantees that $x_\mu z_T = x_v(x_\mu z_\Psi) + [x_\mu, x_v] z_\Psi$, and we know $x_\mu z_\Psi = z_\mu z_\Psi + w$ for some $w \in S_{m-2}$. Since $v \leq \Psi$ and $v \leq \mu$, (C_m) already applies to $x_\lambda(x_v(z_\mu z_\Psi))$. By induction, we know (C_m) also applies to $x_\lambda(x_v w)$, and thus to $x_\lambda(x_v(x_\mu z_\Psi))$. Consequently,

$$x_\lambda(x_\mu z_T) = x_v(x_\lambda(x_\mu z_\Psi)) + [x_\lambda, x_v](x_\mu z_\Psi) + [x_\mu, x_v](x_\lambda z_\Psi) + [x_\lambda, [x_\mu, x_v]] z_\Psi.$$

Recall that λ and μ are interchangeable throughout this argument. If we interchange them in the above equation, and subtract the two, we obtain $x_\lambda(x_\mu z_T) - x_\mu(x_\lambda z_T)$, which is equivalent to

$$x_v(x_\lambda(x_\mu z_\Psi)) - x_v(x_\mu(x_\mu z_T)) + [x_\lambda,[x_\mu,x_v]]z_\Psi - [x_\mu,[x_\lambda,x_v]]z_\Psi.$$

But this is the same as

$$x_v([x_\lambda,x_\mu]z_\Psi) + [x_\lambda,[x_\mu,x_v]]z_\Psi + [x_\mu,[x_v,x_\lambda]]z_\Psi,$$

which can be written as

$$[x_\lambda,x_\mu](x_v z_\Psi) + ([x_v,[x_\lambda,x_\mu]] + [x_\lambda,[x_\mu,x_v]] + [x_\mu,[x_v,x_\lambda]])z_\Psi.$$

And thanks to the Jacobi identity, the terms in parenthesis vanish, and we obtain $[x_\lambda,x_\mu]z_\Psi$. This proves (C_m) and with it the lemma. ■

Great! Just two more to go!

Lemma 6.10. *There exists a representation $\rho : \mathfrak{g} \to \mathfrak{gl}(\mathrm{Sym})$ satisfying:*

1. *$\rho(x_\lambda)z_\Sigma = z_\lambda z_\Sigma$ for $\lambda \leq \Sigma$,*
2. *$\rho(x_\lambda)z_\Sigma \equiv z_\lambda z_\Sigma$ (mod S_m), if Σ has length m.*

Proof. Lemma 6.9 allows us to define a linear map $f : \mathfrak{g} \otimes \mathrm{Sym} \to \mathrm{Sym}$ satisfying (A_m), (B_m), and (C_m) for all m, since f_m restricted to $\mathfrak{g} \otimes S_{m-1}$ coincides with f_{m-1} by the uniqueness part. In other words, Sym becomes a $\mathfrak{g}$-module by condition (C_m), giving us a representation ρ satisfying the conditions 1 and 2 listed above, thanks to conditions (A_m) and (B_m). ■

Lemma 6.11. *Let $t \in T_m \cap J$, where $J = \mathrm{Ker}(\pi)$, and π is the quotient map from T to $\mathfrak{Ug}$. Then the homogenous component t_m of t of degree m lies in I, the kernel of the quotient map taking T to* Sym.

Proof. Write t_m as a linear combination of basis elements $x_{\Sigma(i)}$ for $1 \leq i \leq r$, and where each $\Sigma(i)$ is of length m. The Lie algebra homomorphism $\rho : \mathfrak{g} \to \mathfrak{gl}(\mathrm{Sym})$ constructed in Lemma 6.10 extends, by the universal property of $\mathfrak{Ug} = T/J$, to an algebra homomorphism $\rho' : T \to \mathrm{End}(\mathrm{Sym})$, with $J \subset \mathrm{Ker}(\rho')$. So $\rho'(t) = 0$. Then

$$\rho'(x_{\Sigma(i)}) \cdot 1 = \rho'(x_{\Sigma(i)_1} \otimes \cdots \otimes x_{\Sigma(i)_m}) \cdot 1 = 0$$

by the definition of $x_{\Sigma(i)}$. But this becomes

$$\rho(x_{\Sigma(i)_1}) \cdots \rho(x_{\Sigma(i)_m}) \cdot 1 = 0$$

due to the fact that ρ' is an algebra homomorphism and because the restriction of ρ' to $\mathfrak{g}$ is ρ. And then by Lemma 6.10, we obtain $z_{\Sigma(i)}$. Hence $\rho'(t) \cdot 1$ is a polynomial whose term of highest degree is the appropriate combination of the $z_{\Sigma(i)}$ $(1 \leq i \leq r)$. Therefore this combination of the $z_{\Sigma(i)}$ is 0 in Sym, and $t_m \in I$ as required. ■

Proof of the Poincaré-Birkhoff-Witt Theorem. Let $t \in T^m$, and $\pi : T \to \mathfrak{Ug}$ be the quotient map. We must show that $\pi(t) \in U_{m-1}$ implies $t \in I$. But $t \in T^m$ and $\pi(t) \in U_{m-1}$ together imply that $\pi(t) = \pi(t')$ for some $t' \in T_{m-1}$, and so $t - t' \in J$. Applying Lemma 6.11 to the tensor $t - t' \in T_m \cap J$, and using the fact that the homogenous component of degree m is t, we get $t \in I$. ■

Acknowledgments

First and foremost, I would like to thank my mentor, Jared Bass, for all of his help this summer—without his guidance, this paper would not have been possible. And of course, I would like to thank Peter May for creating and running the REU.

References

[1] Brian Hall. *Lie Groups, Lie Algebras, and Representations.* Springer-Verlag New York Inc. 1st Edition. 2003.

[2] Jeffrey Lee. *Manifolds and Differential Geometry.* American Mathematical Society. 1st Edition. 2009.

[3] Alexander Kirillov, Jr. *An Introduction to Lie Groups and Lie Algebras.* Cambridge University Press. 1st Edition. 2008.

[4] James Humphreys. *Introduction to Lie Algebras and Representation Theory.* Springer-Verlag New York Inc. 3rd Edition. 1980.