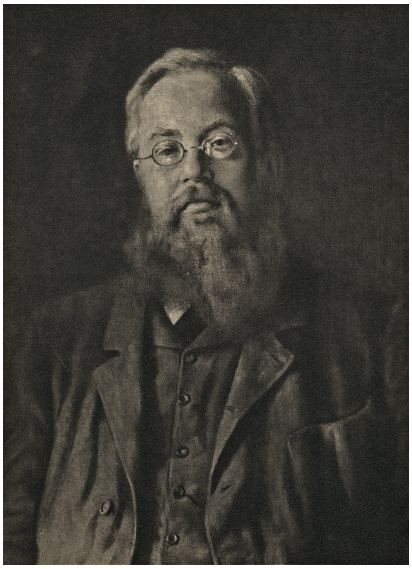
Notes on Lie Theory

Markus Szymik

v0.20201118101340

These are my notes related to the course MA3407, Introduction to Lie Theory, taught at NTNU Trondheim in Spring 2019 and Fall 2020.

Please send comments or questions to markus.szymik@ntnu.no. Do not forget to use your student account, @stud.ntnu.no, account for this.



Marius Sophus Lie (17 December 1842 – 18 February 1899)

Symmetry

Lie theory is about mathematical structures that describe symmetries and allow us to study them.

> *Symmetry is a vast subject, significant in art and nature. Mathematics lies at its root, and it would be hard to find a better one on which to demonstrate the working of the mathematical intellect.* [Wey52]

Symmetries are often–but not exclusively, as we shall see–studied in terms of an algebraic structure: groups.

> *Next to the concept of a function, which is the most important concept pervading the whole of mathematics, the concept of a group is of the greatest significance in the various branches of mathematics and its applications.* [Ale59]

Groups can be very complicated. As often in mathematics: when we meet a complicated problem, we try to approximate it using linear algebra. For instance, this is one of the basic principles of differential calculus. Here, this paradigm is one of the reasons why Lie algebras enter the study of symmetries. They describe symmetries 'infinitesimally.'

Many different situations in mathematics naturally lead to Lie algebras: Lie groups, of course, but also discrete and algebraic groups, as well as others. One might be inclined to argue that Lie groups and algebraic groups, such as the general linear matrix groups, are closer to the core subject than discrete groups, and I would agree. Nevertheless, the former require a non-negligible amount of geometry and topology even to understand their definition. For didactical reasons, this leads us to start with the pure algebra of discrete groups.

Part I: Algebra

In this part, we will deal with the algebraic part of Lie theory concerned with Lie algebras. We think of this theory as an infinitesimal version of the idea of symmetries and groups. Therefore, in Section 1, we start by defining groups to ensure that we are all on the same page. We shall see a definition of groups that is easily transportable to other contexts. We don?t only have to consider groups in the category of sets; we can also look at the group objects in topological spaces, in manifolds, or affine schemes. There will be an occasion to do so later in this course. Particular emphasis was on introducing operations on groups, not only multiplication but also, for instance, conjugation and commutators. This point of view leads to one interpretation of the free group on r generators as the set of all operations on groups of arity r, in Section 2. In Section 3, we look at commutators in more detail and study identities satisfied by them, for instance, the Jacoby identity, which only holds up to conjugation and the Hall–Witt formula. As a generalization of abelianization, we will then define the lower central series in Section 4. We can then pass to the associated graded abelian group and notice that it is equipped naturally with a Lie algebra structure. This construction motivates the general definition of Lie algebras in Section 5. In Section 6, we will consider derivations, which are an infinitesimal version of automorphisms.

From Section 7 on, we study the relationship between Lie algebras and associative algebras. There's a forgetful functor from associative algebras to Lie algebras, using another commutator construction different from that in groups. Some of the most important Lie algebras, those of endomorphisms of vector spaces, arise this way. They are the starting point for the representation theory of Lie algebras. The forgetful functor also has a left-adjoint. This functor sends a Lie algebra to its universal enveloping algebra, and we can construct it as a quotient of the tensor algebra. In Section 8, we prove the Poincaré–Birkhoff–Witt theorem, which describes the associated graded of the universal envelope algebra. It turns out that this is always a symmetric algebra, and we can think of its elements as polynomials. In particular, the universal enveloping algebra and the symmetric algebra have the

same size. We also see that (most) Lie algebras embed into the universal enveloping algebras.

In the following Section 9, we introduce Hopf algebras as a concept that shall play an important role later on in this course. Universal enveloping algebras of Lie algebras are examples of Hopf algebras, as we shall see in Section 10. We will also see how we can, in most cases, recover a Lie algebra inside its universal enveloping algebra as the subset of primitive elements.

For formal reasons, the universal enveloping algebra of a free Lie algebra is the tensor algebra. This fact allows us, in Section 11 to describe the size of the free Lie algebra in a formula that involved the Möbius function. Cool! In Section 12, we will apply this to show that the associated graded of a free group is a free Lie algebra, which is not formal.

Section 13, the final full section of this part, introduces exponentials and the Baker–Campbell–Hausdorff formula. Some related algebraic structures are mentioned briefly in Section 14.

1 Groups

This section has several purposes. First, we need to review the some of the foundations of group theory so that we are all on the same page. Second, we will have occasion to learn about several other algebraic theories in this course, and it is a good idea to think about some general principles in the basic and already somewhat familiar example of groups before we move on to other theories. Third, we can already see one definition of a Lie group and other, more geometrically structured groups.

From one point of view, an algebraic structure on a set X is given by a set of generating n–ary operations

$$X^n \longrightarrow X, \tag{1.1}$$

where $n = 0, 1, 2, \dots$ will depend on the operation. These data are usually required to satisfy another set of relations, and these relations say that other operations (1.1), which are formed from the defining operations and projections onto factors, are equal. We leave it to the general algebraists [Ber15] to formalize this; for our purposes it is enough to know that maps of the form (1.1) are the most relevant for our purposes, and for the start, we shall build our algebraic theories on them.

Observe that the generating operations and relations are by no means unique. There is always a choice in how to describe a given theory, and there is little point in trying to make the most efficient choice. We usually go for an option that is easy to motivate and/or easy to remember.

1.1 A set of defining operations for the theory of groups

Definition 1.1. A *group* is a set G together with three maps

$$G \times G \longrightarrow G,\ (g,h) \longmapsto g \cdot h$$
$$G \longrightarrow G,\ g \longmapsto g^{-1}$$
$$\star \longrightarrow G,\ \star \longmapsto e$$

such that the equalities

$$f(gh) = (fg)h \tag{1.2}$$
$$ge = g = eg \tag{1.3}$$
$$gg^{-1} = e = g^{-1}g \tag{1.4}$$

hold for the indicated maps $G^n \to G$, for suitable n.

Groups form a category together with the following, obvious, definition of morphisms between them. We refer to Appendix A for a summary of such jargon.

Definition 1.2. A *morphism* $G \to H$ of groups is a map that is compatible with the three operations.

Note that the axioms do not play a role.

Exercise 1. The singleton $\star$ has a unique group structure. Show that a constant map $\star \to G$ into a group G is a morphism of groups if and only if the image is the unit e of G.

Exercise 2. A group G is *abelian* if and only if $gh = hg$ for all g, h in G. Show that a group G is abelian if and only if the inversion $G \to G$ is a morphism of groups.

1.2 More operations

The three operations that appear in the definition of a group (the binary *multiplication*, the unary *inversion*, and the constant *unit*) are not the only operations that

we have in a group. For instance we also have the operations

$$[g,h] = ghg^{-1}h^{-1}$$
$$g \rhd h = ghg^{-1},$$
$$\langle f,g,h\rangle = fg^{-1}h$$

in any group. These are the *commutator*, *conjugation*, and the *heap* (or *Malcev*) operation.

The commutator will be studied soon enough, in the following Section 3. Let us first play a bit with the other ones for practice.

Exercise 3. Let g and h be elements in a group. Show that gh and hg are conjugate: there is an element c in G such that $c \rhd gh = hg$.

Exercise 4. Show that the equations

$$f \rhd (g \rhd h) = (f \rhd g) \rhd (f \rhd h)$$

and

$$g \rhd g = g$$

hold for all elements f, g, h in a group G. This exercise verifies that every group has an underlying *quandle* structure. Quandles, in themselves, are an interesting algebraic structure that, among other things, is very useful in knot theory. If we omit the last condition $g \rhd g = g$, we arrive at the notion of a *rack*; these are sets together with a binary operation $\rhd$ such that all left-multiplications $y \mapsto x \rhd y$ are automorphisms. We'll meet the Lie analog of racks in Section 14.1.

Exercise 5. Show that the equations

$$\langle g,h,\langle i,j,k\rangle\rangle = \langle g,\langle j,i,h\rangle,k\rangle = \langle\langle g,h,i\rangle,j,k\rangle$$

and

$$\langle g,g,h\rangle = h, \quad \langle g,h,h\rangle = g$$

hold for all elements $g, \dots$ in a group G. A set X together with a ternary operation $(f,g,h) \mapsto \langle f,g,h\rangle$ that satisfies the axioms in this exercise is called a *heap*.

Exercise 6. Show that the set of bijections $A \to B$ between two sets A and B is a heap with respect to the operation $\langle f, g, h\rangle = fg^{-1}h$. Note the case when A and B do not have the same cardinality.

Exercise 7. It is clear from Exercise 5 that every group G defines a pointed heap (i.e. a heap together with a distinguished element). Show that, conversely, a pointed heap is the same thing as a group–the distinguished element is the unit. Hint: we have to definite the multiplications and the inversion, and verify the axioms.

Needless to say, we still haven't seen all operations in a group:

Exercise 8. Show that a group G is abelian if and only if the *squaring operation*

$$G \longrightarrow G,\ g \longmapsto g^2$$

is a morphism of groups.

1.3 A glimpse ahead: groups in other contexts

Definition 1.1 has the advantage that it is easily transportable into other, more geometric contexts. A suitable context is given by any category with finite products.

Definition 1.3. A *topological group* is a topological space G together with three continuous maps

$$\begin{aligned} m\colon G \times G &\longrightarrow G,\\ i\colon G &\longrightarrow G,\\ e\colon \star &\longrightarrow G, \end{aligned}$$

such that the axioms (1.2), (1.3), and (1.4) hold for the indicated continuous maps $G^n \to G$, for suitable n.

Let us spell out what these indicated maps are and how we can thereby translate the equations into commutative diagrams. Equation (1.2) becomes the diagram

$$\begin{array}{ccc} G \times G \times G & \xrightarrow{m \times \mathrm{id}} & G \times G \\ {\scriptstyle \mathrm{id} \times m}\downarrow & & \downarrow{\scriptstyle m} \\ G \times G & \xrightarrow[m]{} & G, \end{array}$$

whereas (1.3) becomes

$$\begin{array}{ccccc} G & \xrightarrow{(\mathrm{id},e)} & G \times G & \xleftarrow{(e,\mathrm{id})} & G \\ & \searrow & \downarrow{\scriptstyle m} & \swarrow & \\ & & G, & & \end{array}$$

and (1.4) becomes

$$\begin{array}{ccccc} G \times G & & \xrightarrow{\mathrm{id} \times i} & & G \times G \\ {\scriptstyle \Delta}\uparrow & & & & \downarrow{\scriptstyle m} \\ G & \longrightarrow & \star & \xrightarrow{e} & G \\ {\scriptstyle \Delta}\downarrow & & & & \uparrow{\scriptstyle m} \\ G \times G & & \xrightarrow[i \times \mathrm{id}]{} & & G \times G, \end{array}$$

where $\Delta = (\mathrm{id}, \mathrm{id})$ is the diagonal.

Definition 1.4. A *Lie group* is a smooth manifold G together with three smooth maps

$$\begin{aligned} m &\colon G \times G \longrightarrow G, \\ i &\colon G \longrightarrow G, \\ e &\colon \star \longrightarrow G, \end{aligned}$$

such that the axioms (1.2), (1.3), and (1.4), in their diagrammatic form, hold for the indicated smooth maps $G^n \to G$, for suitable n.

Of course, these definitions are only helpful once smooth manifolds and smooth maps are familiar terms. The point here is that having this knowledge is now the *main* difficulty; the rest of the definition is just formally transported from group theory.

Definition 1.5. An *affine algebraic group* over a field K is an affine K–scheme G together with three K–morphisms

$$\begin{aligned} G \times G &\longrightarrow G,\ (g,h) \longmapsto g \cdot h \\ G &\longrightarrow G,\ g \longmapsto g^{-1} \\ \star &\longrightarrow G,\ \star \longmapsto e \end{aligned}$$

such that the axioms (1.2), (1.3), and (1.4), in their diagrammatic form, hold for the indicated K–morphisms $G^n \to G$, for suitable n.

Again, this definition is only meaningful for those who know what an affine K–scheme is. Luckily, that is *not* difficult to say at all, much easier than the definition of a smooth manifold: the category of affine K–schemes is just the opposite of the category of commutative K–algebras with unit. The product of affine K–schemes corresponds to the tensor product, over K, of commutative K–algebras with unit. We can therefore rewrite the definition of an affine algebraic group in terms of a commutative K–algebra H with unit and three morphisms

$$\begin{aligned} \mu &\colon H \longrightarrow H \otimes H \\ \iota &\colon H \longrightarrow H \\ \varepsilon &\colon H \longrightarrow K. \end{aligned}$$

This leads straightforwardly to the notion of a Hopf algebra, which we will discuss later, in Section 9. We'll come back to affine group schemes in Section 21.

Exercise 9. Let us say, for the sake of this exercise, that a *groupgroup* is a group G together with three morphisms

$$\begin{aligned} m &\colon G \times G \longrightarrow G, \\ i &\colon G \longrightarrow G, \\ e &\colon \star \longrightarrow G, \end{aligned}$$

of groups, such that the axioms (1.2), (1.3), and (1.4) hold for the indicated morphisms $G^n \to G$ of groups, for suitable n. Show that a groupgroup is the same

thing as an abelian group. Hint: a groupgroup, by definition, comes with *two* multiplications, *two* inverses, and *two* units, and we can first show that these agree and then that they are abelian. It's a lengthy endeavor, but worth it.

2 Free groups

The contents of this section are a litte more abstract. The goal is to formalize the concept of an 'operation' on groups, so that it can be explored further. It will naturally lead the one explicit model of the free groups.

2.1 Free groups

Definition 2.1. A *free group on r generators* is a group F_r together with r elements $x_1, \ldots, x_r \in \mathrm{F}_r$ such that the natural map

$$\mathrm{Mor}(\mathrm{F}_r, G) \cong G^r,\ \alpha \mapsto (\alpha(x_1), \ldots, \alpha(x_r))$$

is a bijection for all groups G.

In other words: a free group F_r represents the functor $U^r \colon G \mapsto G^r$ from the category of groups to the category of sets. The choice of the x_j corresponds to the choice of a natural isomorphism $U^r \cong \mathrm{Mor}(\mathrm{F}_r, ?)$ of functors.

Example 2.2. The trivial group $\{e\}$ is a free group on zero generators.

Example 2.3. The infinite cyclic group $\mathbb{Z}$ is a free group on one generator.

Proposition 2.4. *For each $r \geqslant 0$, a free group F_r on r generators exists.*

Proof. Let I be a set (!) of groups such that every countable group is isomorphic to a group in I. Let J be the set (!) of all maps $j \colon \{1, \ldots, r\} \to G$ to groups G in I. We'll write G_j for the target of j.

Consider the product group

$$P = \prod_{j \in J} G_j.$$

Note that most of the groups in the set I will appear as factors in the group P many times. For each $k \in \{1, \dots, r\}$, the family $x_k = (j(k) \,|\, j \in J)$ defines an element in P_k.

The group P has the following property: for each group G and each $(g_1, \dots, g_n)$ there is at least on morphism $P \to G$ that takes x_k to g_k for all k. To see that, note that the $g_1, \dots, g_n$ generate a finitely generated subgroup G. Hence, there is a group H in I that is isomorphic to it, and a choice of isomorphism defines a map $j\colon \{1, \dots, r\} \to H = G_j$ such that the embedding $G_j \to G$ takes $j(k)$ to g_k for all k. Then the composite

$$P \xrightarrow{\mathrm{pr}_j} G_j \longrightarrow G$$

with the projection is a map as desired.

The morphism $P \to G$ is not unique, but it is uniquely determined on the subgroup $\mathrm{Sp}(x_1, \dots, x_k)$ generated by the x_k, so that this subgroup is free on the x_k. □

Exercise 10. Let P the group defined in the proof above, and let $\mathrm{Eq}(x_1, \dots, x_k)$ be the subgroup that consists of all elements $y \in P$ that are fixed by all morphisms $\alpha\colon P \to P$ that fix all the x_k. So, in particular, all the elements x_k lie in this subgroup, and we, therefore, have $\mathrm{Sp}(x_1, \dots, x_k) \leqslant \mathrm{Eq}(x_1, \dots, x_k)$. Show that the subgroup $\mathrm{Eq}(x_1, \dots, x_k)$ is also free on the x_k. After all we already know, this is equivalent to $\mathrm{Eq}(x_1, \dots, x_k) = \mathrm{Sp}(x_1, \dots, x_k)$.

2.2 All operations

Definition 2.5. An *n–ary operation for groups* is a family

$$\Phi_G\colon G^n \longrightarrow G$$

of maps, not necessarily morphisms of groups, one for each group G, that is compatible with all morphisms between groups. This means that

$$\Phi_H(\alpha(g_1), \dots, \alpha(g_n)) = \alpha(\Phi_G(g_1, \dots, g_n)) \tag{2.1}$$

for all morphisms $\alpha\colon G \to H$ and all $(g_1, \ldots, g_n)$ in G^n.

The equation can be translated into the commutative diagram

$$\begin{array}{ccc} G^n & \xrightarrow{\Phi_G} & G \\ {\scriptstyle \alpha^n}\downarrow & & \downarrow{\scriptstyle \alpha} \\ H^n & \xrightarrow[\Phi_H]{} & H \end{array}$$

of maps of sets. In other words, an n–ary operation $\Phi = (\Phi_G \,|\, G \text{ group})$ for groups is a natural transformation

$$U^n \longrightarrow U,$$

where U is the forgetful functor from the category of groups to the category of sets.

Example 2.6. The projections

$$\Pi_j\colon G^n \longrightarrow G,\ (g_1, \ldots, g_n) \longmapsto g_j \tag{2.2}$$

for $j = 1, \ldots, n$ are n–ary operation for groups.

The functor U^n is represented by the free group F_n on n generators. Therefore, the Yoneda lemma shows that there is a bijection between the sets of n–ary operation for groups and the underlying set of the free group F_n on n generators.

Proposition 2.7. *There is a bijection between a free group on r generators and the set of n–ary operation for groups.*

It is not clear at all from the definition that there is only a set of n–ary operations because there is definitely *not* a set of groups.

We can turn these abstract arguments around and *define* the free group on n generators as the sets of n–ary operations for groups together with the following group structure: The product of two elements Φ and Ψ is defined as $(\Phi \cdot \Psi)_G = \Phi_G \cdot \Psi_G$, which is the composition

$$\Phi_G \cdot \Psi_G\colon G^n \xrightarrow{(\Phi_G, \Psi_G)} G \times G \xrightarrow{\quad m \quad} G.$$

Similarly, the inverse is defined by

$$\Phi_G^{-1}\colon G^n \xrightarrow{\Phi_G} G \xrightarrow{i} G$$

on G, and the unit is defined by

$$G^n \longrightarrow \star \xrightarrow{e} G$$

on G. It is easy to check that these operations turn the set F_n of n–ary operations for groups into a group such that the bijections in Proposition 2.7 are isomorphisms.

Remark 2.8. Given a morphism $\alpha\colon \mathrm{F}_n \to G$ of groups, we can evaluate it on the n projections (2.2) to get n elements in G:

$$\alpha(\Phi_j) \in G.$$

Conversely, given $(g_1, \ldots, g_n) \in G^n$, we can define a morphism $\mathrm{F}_n \to G$ by sending an n–ary operation Φ for groups to the image of $(g_1, \ldots, g_n) \in G^n$ in G under Φ_G. These maps are inverse to each other. However, it is not clear how to verify the equation

$$\Phi = \Phi_{\mathrm{F}_r}(\Pi_1, \ldots, \Pi_r)$$

without assuming the existence of free groups. In any event, we see that the projections (2.2) are the generators of the free group in this model.

2.3 Exercises

Exercise 11. Describe all 0–ary (constant) operations for groups.

Exercise 12. Describe all 1–ary (unary) operations for groups.

3 Commutators

Recall the definition of the commutator

$$[g, h] = ghg^{-1}h^{-1}$$

of two elements g, h in a group G from Section 1.2. We have $[x, y] = e$ if and only if x and y commute ($yx = xy$).

We have

$$[x, x] = e \tag{3.1}$$

and more generally

$$[x^m, x^n] = e$$

for all integers m and n. In particular,

$$[x, e] = e.$$

We have

$$[y, x] = [x, y]^{-1}, \tag{3.2}$$

so that the inverse of a commutator is again a commutator, and

$$x \rhd [y, z] = [x \rhd y, x \rhd z], \tag{3.3}$$

because conjugation is a morphism of groups, so that any conjugate of a commutator is also a commutator.

We have

$$[x, y] = xyx^{-1}y^{-1} = (x \rhd y)y^{-1},$$

and this implies

$$x \rhd y = [x, y]y. \tag{3.4}$$

In other words: conjugation is left multiplication with the commutator. In the case when y itself is a commutator, we get

$$x \rhd [y, z] = [x, [y, z]][y, z] \tag{3.5}$$

The operation $y \mapsto [x, y]$ is not quite a morphism of groups:

Proposition 3.1.

$$[x, yz] = [x, y](y \rhd [x, z])$$
$$[xy, z] = (x \rhd [y, z])[x, z]$$

Proof. As for the first equation, we could try to be clever here, but it is faster if we just expand the right hand side and see what we get:

$$[x, y](y \rhd [x, z]) = xyx^{-1}y^{-1}yxzx^{-1}z^{-1}y^{-1}.$$

Then we can cancel the term $x^{-1}y^{-1}yx$ in the middle to get $xyzx^{-1}z^{-1}y^{-1}$, and this is the left hand side $[x, yz]$. The argument for the second equation is similar. □

Exercise 13. Spell out a similar argument for the second equation. Then try to be clever: is there a more insightful way to explain the result?

Exercise 14. Prove the equation

$$[x, yz][y, zx][z, xy] = e$$

for all elements x, y, z in any group.

Theorem 3.2. (Hall–Witt)

$$[z \rhd x, [y, z]][y \rhd z, [x, y]][x \rhd y, [z, x]] = e$$

Proof. Again, it is straightforward to verify this. First, we have

$$[z \rhd x, [y, z]] = zxz^{-1}yzy^{-1}z^{-1}zx^{-1}z^{-1}zyz^{-1}y^{-1},$$

and we can cancel this to get

$$[z \rhd x, [y, z]] = zxz^{-1}yzy^{-1}x^{-1}yz^{-1}y^{-1}. \tag{3.6}$$

Cyclically permuting the elements x, y, z we get:

$$[y \rhd z, [x, y]] = yzy^{-1}xyx^{-1}z^{-1}xy^{-1}x^{-1}, \tag{3.7}$$
$$[x \rhd y, [z, x]] = xyx^{-1}zxz^{-1}y^{-1}zx^{-1}z^{-1}. \tag{3.8}$$

Then we multiply (3.6), (3.7), and (3.8) to see that the product is e. □

Recall that $[x,y][y,x] = e$. The Hall–Witt identity can be thought of as a three-variable analog of this equation. Cohen's paper [Coh17] develops this idea further for more variables.

Exercise 15. We can write the equation $[x,y] = e$ in the form $xy = yx$. Can we write the equation $[x,[y,z]] = e$ in the form $v = w$, where v and w are words involving x, y, and z, but *not* involving their inverses?

3.1 Products of commutators

We have seen in (3.2) that the inverse of a commutator is again a commutator, and the neutral element e in any group is obviously also a commutator (3.1). In contrast, the product of commutators need not be a commutator. The following example seems fitting in a course on Lie theory.

Example 3.3. Let $\mathrm{SL}_2(\mathbb{R})$ be the group of real $(2,2)$–matrices with determinant 1. Then the element

$$-E = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

is a product of commutators, but not a commutator itself.

Let's first see that it is a product of commutators. One way of doing this is to start with the equation

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}.$$

The matrices on the right hand side are commutators because of the following identities:

$$\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} = \left[\begin{pmatrix} \sqrt{2} & 0 \\ 0 & \frac{1}{2}\sqrt{2} \end{pmatrix}, \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \right]$$
$$\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix} = \left[\begin{pmatrix} \frac{1}{2}\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix} \right].$$

Let us now see that $-E$ is not a commutator. Assume it were, so that we have an equation $-E = XYX^{-1}Y^{-1}$ or $XYX^{-1} = -Y$. Computing the traces on both sides, we get

$$\operatorname{tr}(Y) = \operatorname{tr}(XYX^{-1}) = \operatorname{tr}(-Y) = -\operatorname{tr}(Y),$$

so that $\operatorname{tr}(Y) = 0$. Since we also know $\det(Y) = 1$, the two complex eigenvalues of Y are $\pm\mathrm{i}$, and Y is conjugate to a rotation with angle $\pi/2$: there is a matrix G in the group $\mathrm{SL}_2(\mathbb{R})$ such that

$$G \rhd Y = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then $-E = G \rhd (-E) = [G \rhd X, G \rhd Y]$. We write

$$G \rhd X = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

and see what we get from the equation $(G \rhd X)(G \rhd Y) = -(G \rhd Y)(G \rhd X)$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} b & -a \\ d & -c \end{pmatrix}$$
$$-\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ -a & -b \end{pmatrix}.$$

This shows $c = b$ and $d = -a$. Then $\det(G \rhd X) = -a^2 - b^2 < 0$ is in contradiction to the assumption $G \rhd X \in \mathrm{SL}_2(\mathbb{R})$.

3.2 Commutator subgroups and abelianization

Let G be a group. If U and V are subgroups then

$$[U, V]$$

denotes the subgroup generated by the commutators $[u, v]$ for $u \in U$ and $v \in V$. It follows from (3.2) that $[V, U] = [U, V]$. If U and V are normal subgroups, their product

$$UV = \{uv \,|\, u \in U, v \in V\}$$

is also a normal subgroup, and it follows from (3.3) that $[U, V]$ is a normal subgroup, too.

The following result will be used further down in the proof of Proposition 4.3.

Lemma 3.4. *If U, V, and W are normal subgroups of G, then*

$$[U, [V, W]] \leqslant [V, [W, U]][W, [U, V]]$$

Proof. The follows from the Hall–Witt identity (see Theorem 3.2). It suffices to show that $[u, c] \in [V, [W, U]][W, [U, V]]$ if $u \in U$ and $c \in [V, W]$ because the elements genedrate the group $[U, [V, W]]$.

We first do the case $c = [v, w]$. Then we write

$$u = w \rhd (w^{-1} \rhd u) = w \rhd u'$$

with $u' = w^{-1} \rhd u$ still in U because U is normal. The Hall–Witt identity gives

$$[u, [v, w]] = [w \rhd u', [v, w]] = [u' \rhd v, [w, u']^{-1}[v \rhd w, [u', v]]^{-1},$$

and this lies in $[V, [W, U]][W, [U, V]]$ because V and W are normal.

In general c is a product of commutators, and we have to show that an element of the form $[u, cc']$ lies in $[V, [W, U]][W, [U, V]]$ if both $[u, c]$ and $[u, c']$ lie in it. By Proposition 3.1, we have

$$[u, cc'] = [u, c](c \rhd [u, c']).$$

Then the result follows because $[V, [W, U]]$ and $[W, [U, V]]$ are normal, and hence so is their product. □

Definition 3.5. The subgroup $[G, G]$ is generated by *all* commutators in a group and is called the *commutator subgroup* of the group G.

It follows from (3.3) that the commutator subgroup is normal in G. By definition, the commutator of any two elements lies in $[G, G]$, so the factor group is abelian.

Definition 3.6. The factor group

$$G^{\mathrm{ab}} = G/[G,G]$$

is called the *abelianization* of G.

Proposition 3.7. *There is a natural bijection*

$$\mathrm{Hom}(G^{\mathrm{ab}}, A) \cong \mathrm{Mor}(G, A)$$

for groups G and abelian groups A.

Proof. The proof is easy and can be left as an exercise: Use that

$$\mathrm{Mor}(G/N, H) \cong \{\varphi\colon G \to H \mid \varphi(N) = e\}$$

and that $\varphi([G,G]) = e$ is automatic if H is abelian. □

In other words, the functor $G \mapsto G^{\mathrm{ab}}$ is left-adjoint to the forgetful functor from the category of abelian groups to the category of all groups.

Exercise 16. There is an alternative construction of the abelianization of a group G: let $\mathbb{Z}G$ be the free abelian group with basis G. Its elements are the linear combinations

$$\sum_{g \in G} \lambda_g \cdot g,$$

with $\lambda_g = 0$ for all but finitely many group elements $g \in G$. The quotient by the subgroup generated by all elements of the form

$$g + h - gh,$$

for $g, h \in G$, is isomorphic to G^{ab}. There are at least two ways of showing this: by direct comparison and by verifying the universal property in Proposition 3.7.

Exercise 17. Check that $\mathrm{F}_r^{\mathrm{ab}} \cong \mathbb{Z}^r$.

4 The lower central series

Definition 4.1. Let G be a group. The *lower central series* is the sequence

$$G = \Gamma_1(G) \geqslant \Gamma_2(G) \geqslant \Gamma_3(G) \geqslant \dots$$

of subgroups that are defined inductively by $\Gamma_1(G) = G$ and

$$\Gamma_{m+1}(G) = [G, \Gamma_m(G)].$$

In particular, we have $\Gamma_2(G) = [G, G]$. A group G is trivial if and only $\Gamma_1(G) = e$, and it is abelian if and only if $\Gamma_2(G) = e$.

Definition 4.2. A group is called *nilpotent* if $\Gamma_m(G) = e$ for some m.

It is true, but not obvious, that a finite group is nilpotent if and only if it is a product of Sylow subgroups. In particular, every p–group is nilpotent.

Infinite groups can have

$$\bigcap_m \Gamma_m(G) = e$$

without being nilpotent. These groups are called *residually nilpotent*. The free groups F_r with $r \geqslant 2$ are examples.

Proposition 4.3. *We have*

$$[\Gamma_m(G), \Gamma_n(G)] \leqslant \Gamma_{m+n}(G).$$

Proof. We can prove this by induction, say on m. For $m = 1$ the equation holds by definition. Let us therefore assume that it holds for all $l < m$. Then we have

$$\begin{aligned}
[\Gamma_m(G), \Gamma_n(G)] &= [[\Gamma_{m-1}(G), G], \Gamma_n(G)] \\
&\leqslant [[G, \Gamma_n(G)], \Gamma_{m-1}(G)][[\Gamma_n(G), \Gamma_{m-1}(G)], G] \\
&\leqslant [\Gamma_{n+1}(G)], \Gamma_{m-1}(G)][\Gamma_{n+m-1}(G), G] \\
&\leqslant \Gamma_{m+n}(G)\, \Gamma_{m+n}(G) \\
&= \Gamma_{m+n}(G)
\end{aligned}$$

by definition, Lemma 3.4, and induction. □

Proposition 4.4. *The subgroups $\Gamma_m(G)$ are normal in G for all m.*

Proof. We can prove this by induction, say on m. For $m = 1$ the equation holds by definition. Otherwise, we know that $\Gamma_m(G) = [G, \Gamma_{m-1}(G)]$ is generated by elements $[g, x]$ where $g \in G$ and $x \in \Gamma_{m-1}(G)$. If we conjugate by and h in G, Equation (3.3) gives

$$h \rhd [g, x] = [g', x']$$

with $g' \in G$ and $x' \in \Gamma_{m-1}(G)$ by induction. Then $h \rhd [g, x] \in \Gamma_m(G)$ follows. □

Proposition 4.5. *The factor groups $\Gamma_m(G)/\Gamma_{m+1}(G)$ are abelian for all*

The case $m = 1$ has already been mentioned before Definition 3.6: the abelianization $G/[G, G]$ is an abelian group.

Proof of Proposition 4.5. Let $x, y \in \Gamma_m(G)$. Then the commutator $[x, y]$ lies in $\Gamma_{2m}(G)$ by Proposition 4.3. Since $m \geqslant 1$, we have the inequality $2m \geqslant m + 1$, and this implies the inclusion $\Gamma_{2m}(G) \leqslant \Gamma_{m+1}(G)$. Then $[x, y] \in \Gamma_{m+1}(G)$ implies the result. □

Lemma 4.6. *If $g \in G$ and $x \in \Gamma_m(G)$, then x and its conjugate $g \rhd x$ represent the same element in the factor group $\Gamma_m(G)/\Gamma_{m+1}(G)$.*

Proof. We know that the equation $gxg^{-1}x^{-1} = e$ holds modulo $\Gamma_{m+1}(G)$ by the previous result. Then $gxg^{-1} = x$ modulo $\Gamma_{m+1}(G)$ follows. □

4.1 Minimality of the lower central series

The following results characterizes the lower central series. It will be useful later, in Section 12.

Proposition 4.7. *If*

$$G = V^1 \geqslant V^2 \geqslant V^3 \geqslant \dots$$

is a filtration with $[V^m, V^n] \leqslant V^{m+n}$*, then* $\Gamma_n(G) \leqslant V^n$ *for all* $n \geqslant 1$*.*

Proof. We can prove this by induction on n. The case $n = 1$ is true by definition because $\Gamma_1(G) = G = V^1$. If $n \geqslant 2$, we have by induction

$$\Gamma_n(G) = [\Gamma_1(G), \Gamma_{n-1}(G)] \leqslant [V^1, V^{n-1}] \leqslant V^n,$$

as required. □

4.2 The associated graded

Definition 4.8. Let G be a group. The *associated graded* with respect to the descending filtration given by the lower central series is

$$\mathrm{Gr}(G) = \bigoplus_{m=1}^{\infty} \Gamma_m(G)/\Gamma_{m+1}(G) = G^{\mathrm{ab}} \oplus \dots.$$

For example, if $G = A$ is abelian, then $\mathrm{Gr}(A) = A \oplus 0 \oplus 0 \oplus \cdots = A$. Once the lower central series quotients are zero, they stay zero. We shall work out some more interesting examples in the following Section 4.3.

Remark 4.9. In this version of the definition, the object $\mathrm{Gr}(G)$ is not actually a *graded* abelian group, only an abelian group. This is a bit unfortunate, and we may have to revise it later. However, it's simple and good enough for now, so we'll keep it for at least a while.

The associated graded $\mathrm{Gr}(G)$ of a group G has some extra structure:

Proposition 4.10. *The map* $\mathrm{Gr}(G) \times \mathrm{Gr}(G) \to \mathrm{Gr}(G)$ *induced by the commutator bracket*

$$(x, y) \longmapsto [x, y]$$

is bilinear. It satisfies

$$[x, x] = 0$$

and

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0.$$

Proof. Bilinearity follows from Proposition 3.1 and Lemma 4.6. The first equation is obvious. The second equation is a consequence of the Hall–Witt identity (see Theorem 3.2). □

In the next section, we will define Lie rings in such a way that the preceding proposition shows that $\mathrm{Gr}(G)$ is an example of such a structure.

4.3 Examples

Let us look at some examples to get an idea of what is going on.

Example 4.11. Let us look at the symmetric group S_3 of order 6. The transpositions (12) and (23) do no commute. In fact, we have

$$(12)(23)(12)(23) = (132).$$

This shows that the commutator subgroup has order at least 3. And it cannot be bigger, because the sign of any commutator is even. This shows that

$$[\mathrm{S}_3, \mathrm{S}_3] = \langle (123) \rangle$$

is cyclic of order 3. Since $[\mathrm{S}_3, [\mathrm{S}_3, \mathrm{S}_3]]$ is generated by

$$[(12), (123)] = (123),$$

we have $\Gamma_m(\mathrm{S}_3) = \langle (123) \rangle$ for all $m \geqslant 2$.

We have seen that the group S_3 is not nilpotent. It is a bit more interesting to look at nilpotent groups:

Example 4.12. Let us look at the dihedral group D_8 of order 8. Let's say that r is the rotation of order 4, with angle $\pi/2$, and s is the reflection. Then r and s do not commute. Instead, we have $srs = r^3$, so that

$$[s, r] = r^2.$$

The subgroup $\langle r^2 \rangle$ is normal, and the factor group is isomorphic to the Klein 4–group $V \cong \mathbb{Z}/2 \times \mathbb{Z}/2$. It follows that

$$[\mathrm{D}_8, \mathrm{D}_8] = \langle r^2 \rangle$$

and

$$[\mathrm{D}_8, [\mathrm{D}_8, \mathrm{D}_8]] = e$$

because r^2 is central. The bracket

$$V \times V \to \Lambda^2 V$$

is non-zero, and there is, up to isomorphism, a unique non-zero bracket.

Example 4.13. Let us look at the quaternion group $\mathrm{Q}_8 = \{\pm 1, \pm \mathrm{i}, \pm \mathrm{j}, \pm \mathrm{k}\}$ of order 8. Then $\mathrm{ji} = -\mathrm{ij}$ shows that i and j do not commute, and $[\mathrm{i}, \mathrm{j}] = -1$. By the same argument as above, we have

$$[\mathrm{Q}_8, \mathrm{Q}_8] = \langle -1 \rangle$$

and

$$[\mathrm{Q}_8, [\mathrm{Q}_8, \mathrm{Q}_8]] = e$$

because -1 is central. So $\mathrm{Gr}(\mathrm{Q}_8) \cong \mathrm{Gr}(\mathrm{D}_8)$.

4.4 Code

Here is some GAP code that can be used to do computations automatically.

First, the following function computes the terms of the lower central series.

```
LowCenSer := function( G, n )
    if n = 1 then
        return G;
    else
        return CommutatorSubgroup( G, LowCenSer( G, n-1 ) );
    fi;
end;
```

Then, we collect the terms of the lower central series until it stabilizes.

```
AssGra := function( G )
    local L,n;
    L:=[];
    n:=1;
    while Size( LowCenSer( G, n ) / LowCenSer( G, n+1 ) ) > 1 do
        Append( L, [LowCenSer( G, n ) / LowCenSer( G, n+1 )] );
        n := n+1;
    od;
    return L;
end;
```

Here are two groups to test this with.

```
G := DihedralGroup( 8 );
G := QuaternionGroup( 8 );
```

The result is easier to read if we let GAP describe the structure of these groups.

```
List( AssGra( G ), Q -> StructureDescription( Q ) );
```

5 Lie algebras

Definition 5.1. A *Lie ring* is an abelian group L together with a bilinear map

$$[\ ,\ ]\colon L \times L \longrightarrow L$$

such that

$$[x, x] = 0 \tag{5.1}$$

and

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0 \tag{5.2}$$

hold for all x, y, z in L. More generally, a *Lie algebra* over a commutative ring R with unit is an R–module L together with an R–bilinear map $[\ ,\ ]$ that satisfies (5.1) and (5.2).

The reader will not miss much if they decide to read 'R–module' as 'abelian group' if $R = \mathbb{Z}$ and 'K–vector space' if $R = K$ is a field.

Remark 5.2. We note that (5.1) implies

$$0 = [x + y, x + y] = [x, x] + [x, y] + [y, x] + [y, y] = [x, y] + [y, x],$$

so that

$$[y, x] = -[x, y] \tag{5.3}$$

for all x, y in L. The converse is not true: $[y, x] = -[x, y]$ for all x, y in L only implies $2[x, x] = 0$, so that $[x, x]$ may be 2–torsion without being zero. This difference will not be essential for us, but at least we should point out that the 'operadic' condition (5.3) is weaker than (5.1).

Remark 5.3. The equation (5.2) is often referred to as the *Jacobi identity*.

5.1 Abelian Lie algebras

Every R–module M is a Lie algebra with $[x, y] = 0$ for all x, y in M. Lie algebras of this form are called *abelian*. Note that, by definition, abelian Lie rings are essentially the same thing as abelian groups.

Exercise 18. Define abelianization as a functor from the category of Lie rings to the category of abelian Lie algebras (or: abelian groups) that is adjoint functor to the inclusion functor from the category of abelian Lie rings (or: abelian groups) into the category of Lie algebras.

5.2 Lie rings from groups

Proposition 5.4. *If G is a discrete group, then the associated graded $\mathrm{Gr}(G)$ for the lower central series, together with the bracket $[\ ,\ ]$ induced by the commutator, is a Lie ring.*

Proof. The statement is a reformulation of Proposition 4.10. □

Exercise 19. What happens if G is an abelian group? Is $\mathrm{Gr}(G)$ automatically abelian, too? What about the converse: are there non-abelian groups G such that $\mathrm{Gr}(G)$ is abelian? We'll find that $\mathrm{Gr}(G) = 0$ is equivalent to G being perfect. By definition, this means $\Gamma_1 = \Gamma_2$. The Lie algebra $\mathrm{Gr}(G)$ is abelian if and only if $\Gamma_2 = \Gamma_3$; this holds for the group S_3, for instance, as we have seen in Example 4.11.

Example 5.5. We have described $\mathrm{Gr}(\mathrm{D}_8)$ and $\mathrm{Gr}(\mathrm{Q}_8)$ in Example 4.12 and 4.12, respectively. They are isomorphic Lie rings, even though D_8 and Q_8 are not isomorphic as groups.

A morphism $f\colon G \to H$ of groups induces a morphism

$$\mathrm{Gr}(f)\colon\ \mathrm{Gr}(G) \longrightarrow \mathrm{Gr}(H)$$

of Lie algebras. In general, if f is an isomorphism, so is $\mathrm{Gr}(f)$. But the converse need no hold: check $\mathrm{S}_3 \to \{\pm 1\}$.

Exercise 20. Is there a morphism $f\colon P \to Q$ between finite p–groups such that $\mathrm{Gr}(f)$ is an isomorphism of Lie algebras, but f is not an isomorphism of groups?

In the previous example, there is no morphism of groups $\mathrm{D}_8 \to \mathrm{Q}_8$ or $\mathrm{Q}_8 \to \mathrm{D}_8$ that would induce an isomorphism between the Lie algebras.

5.3 More examples

We shall compute $\mathrm{Gr}(\mathrm{F}_r)$ for the free group F_r on r generators later, in Section 12. Here, we can already look at some simpler pieces. We have already seen that

$$\mathrm{F}_r^{\mathrm{ab}} = \mathrm{F}_r/\Gamma_2(\mathrm{F}_r) \cong \mathbb{Z}^r$$

is (free) abelian. What can we say about the next case, the groups

$$\mathrm{F}_r/\Gamma_3(\mathrm{F}_r)$$

that are nilpotent of class two?

Exercise 21. Show that $\mathrm{F}_2/\Gamma_3(\mathrm{F}_2)$ is isomorphic to the group of all matrices

$$\begin{pmatrix} 1 & x_1 & y \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{pmatrix}$$

with $x_1, x_2, y \in \mathbb{Z}$.

Of course, we also know that $\mathrm{F}_1/\Gamma_3(\mathrm{F}_1) \cong \mathbb{Z}$ is isomorphic to the the group of all matrices

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

with $a \in \mathbb{Z}$, but that is not a start of a pattern:

Exercise 22. Show that $\mathrm{F}_3/\Gamma_3(\mathrm{F}_3)$ is *not* isomorphic to the group of all matrices

$$\begin{pmatrix} 1 & a & b & c \\ 0 & 1 & d & e \\ 0 & 0 & 1 & f \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with $a, \ldots, f \in \mathbb{Z}$. Hint: show that the matrix group is not nilpotent of class 2.

The Lie algebra $\mathrm{Gr}(\mathbb{Z}^r) \cong \mathbb{Z}^r$ is very easy to describe, because the group is abelian. The Lie algebra $\mathrm{Gr}(\mathrm{F}_r/\Gamma_2(\mathrm{F}_r))$ is not much harder: If $x_1, \ldots, x_r$ are the generators of F_r, then $\mathrm{Gr}(\mathrm{F}_r/\Gamma_2(\mathrm{F}_r))$ is free as an abelian group with basis

$$x_i \text{ for } 1 \leqslant i \leqslant r$$
$$[x_j, x_k] \text{ for } 1 \leqslant j < k \leqslant r,$$

and the Lie bracket is as suggested by the notation or zero.

5.4 More exercises

Exercise 23. Show that every 1–dimensional Lie algebra over a field is abelian.

Exercise 24. Classify 2–dimensional Lie algebras over a field.

Exercise 25. Show that the vector product $(v, w) \mapsto v \times w$ equips $\mathbb{R}^3$ with a Lie algebra structure.

Exercise 26. Show that the Lie algebra $\mathbb{R}^3$ from the preceding exercise is *not* isomorphic to $\mathrm{sl}_2(\mathbb{R})$ as real Lie algebras. (Hint: in $\mathbb{R}^3$ we have that $[x, y]$ is zero or linearly independent of x and y; in $\mathrm{sl}_2(\mathbb{R})$ we have the elements H and X with $[H, X] = 2X$.) What happens if we extend scalars to $\mathbb{C}$: is $\mathrm{sl}_2(\mathbb{C})$ isomorphic to $\mathbb{C}^3$ with the same formulas as the vector product?

Exercise 27. Is it true that we always have

$$[w, [x, [y, z]]] + [x, [w, [z, y]]] + [y, [z, [w, x]]] + [z, [y, [x, w]]] = 0$$

in a Lie algebra? We'll later see a systematic approach to such questions.

6 Derivations

In this section, we dwell a bit on the notation of a derivation, and we shall see how that leads to algebraic structures that are closely related to Lie algebras.

For the moment, let us agree that an *algebra* X is an abelian group (or more generally: an R–module for some commutative ring R with unit) together with a bilinear map

$$\star\colon X \times X \longrightarrow X,\ (x,y) \longmapsto x \star y.$$

No further axioms are required. Examples are given by Lie algebras, with

$$x \star y = [x,y]$$

the Lie bracket, and associative algebras (with or without unit), with

$$x \star y = x \cdot y$$

the product.

6.1 Lie algebras from associative algebras

The following observation is fundamental.

Proposition 6.1. *For any associative R–algebra A, with or without unit, the map*

$$[x,y] = xy - yx \tag{6.1}$$

defines a Lie algebra structure on A.

Proof. Bilinearity and the vanishing of $[x,x]$ are obvious. As for the Jacobi identity, write

$$[x,[y,z]] = x(yz-zy) - (yz-zy)x = xyz - xzy - yzx + zyx,$$

and similarly for the other two terms by cyclic permutation. Then observe that the signs in the twelve terms cancel each other. □

We shall denote this Lie algebra by A_{Lie} for clarity.

Example 6.2. The Lie ring A_{Lie} is abelian if and only if A is a commutative ring.

Example 6.3. A very important example of a Lie algebra over a commutative ring R is the Lie algebra that comes from the matrix algebra $\mathrm{M}_n(R)$ of all (n, n)–matrices with entries in R. We shall denote this Lie algebra by

$$\mathrm{gl}_n(R) = \mathrm{M}_n(R)_{\mathrm{Lie}}.$$

More generally, we have the endomorphism Lie algebra

$$\mathrm{gl}(V) = \mathrm{End}(V)_{\mathrm{Lie}}$$

of every R–module V.

The functor $A \mapsto A_{\mathrm{Lie}}$ has an adjoint; we shall discuss it in Section 7.

Remark 6.4. It is possible to ask, more generally, the question whether we can relax the associativity condition on a product so that the bracket (6.1) still gives rise to a Lie algebra. This leads to the notion of a *pre-Lie algebra* (or *Vinberg algebra*). Pre-Lie algebras have been introduced by Gerstenhaber, in his work on deformations of algebras and Hochschild cohomology, and at about the same time by Vinberg in his work on differential geometry. Useful references are the papers by Chapoton–Livernet [CL01], Cartier [Car10], and Fløystad–Munthe-Kaas [FM-K18].

6.2 Automorphisms of groups

Let us recall a few familiar facts about groups. Let G be a group. For any group element g, conjugation $x \mapsto g \rhd x = gxg^{-1}$ with g defines an automorphism of G, and this map $G \to \mathrm{Aut}(G)$ is a morphism of groups:

$$g \rhd (h \rhd x) = (gh) \rhd x.$$

The image is the normal (!) subgroup $\mathrm{Inn}(G)$ of *inner automorphisms* of G. The quotient $\mathrm{Out}(G) = \mathrm{Aut}(G)/\mathrm{Inn}(G)$ is the *outer automorphism group* of G. The kernel of the morphism $G \to \mathrm{Aut}(G)$ is the *center* $\mathrm{Z}(G)$ of G. We can summarize the situation in the sequence

$$1 \longrightarrow \mathrm{Z}(G) \longrightarrow G \longrightarrow \mathrm{Aut}(G) \longrightarrow \mathrm{Out}(G) \longrightarrow 1,$$

which is exact in the sense that the kernels and images agree at each point.

Exercise 28. Make sure that we can verify the facts quoted above.

6.3 Derivations on algebras in general

What are the analogues of those results for Lie algebras?

Definition 6.5. Let A be an R–algebra of sorts, say with a R–bilinear multiplication $\star$ that may or may not verify certain axioms. A *derivation* on $(A, \star)$ is an R–linear map $D\colon A \to A$ such that

$$D(x \star y) = D(x) \star y + x \star D(y)$$

for all x and y in A.

Example 6.6. If the algebra is trivial, with $x \star y = 0$ for all x and y, then all R–linear maps are derivations.

The composite of two derivations is not necessarily another derivation, but their commutator (6.1) is:

Proposition 6.7. *The derivations on* $(A, \star)$ *form a Lie subalgebra* $\mathrm{Der}(A, \star) \leqslant \mathrm{gl}(A)$.

We think of the Lie algebra $\mathrm{gl}(A)$ as a linear approximation to the group $\mathrm{GL}(A)$. Note that neither of them depends on an algebra structure $\star$ on A. If we want

to take such an algebra structure into account, then we think of the Lie algebra $\mathrm{Der}(A, \star)$ as a linear approximation to the automorphism group $\mathrm{Aut}(A, \star)$. We will make this precise later, in Example 23.12.

Here is a standard example, one where the $(A, \star)$ is algebra associative is with unit, and even commutative, because it is an algebra of functions.

Example 6.8. As usual, let $A = R[x_1, \ldots, x_n]$ denote the (commutative) polynomial algebra. Then the Lie algebra $\mathrm{Der}(A, \cdot)$ is an A–module, and as such it is free with basis elements $\partial_j = \frac{\partial}{\partial x_j}$. For a derivation is obviously determined by the values on the generators x_j, and the map $\mathrm{Der}(A) \to A^n$ that sends D to $(D(x_1), \ldots, D(x_n))$ is injective. It is also surjective: given $f = (f_1, \ldots, f_n) \in A^n$, the derivation

$$D_f = \sum_j f_j \partial_j$$

sends x_j to f_j.

Exercise 29. What is the formula for the bracket $[D_f, D_g]$ in this example? We'll come back to this later in Example 6.12.

6.4 Derivations on Lie algebras

Let us now consider the case when $(L, \star)$ is a Lie algebra, with $x \star y = [x, y]$ the Lie bracket. Then $\mathrm{Der}(L) = \mathrm{Der}(L, [\,,\,]) \leqslant \mathrm{gl}(L)$ is a Lie subalgebra.

The Jacobi identity for Lie algebras states that the action of any element on the algebra is a derivation. Writing

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

this can also be rewritten in the following form.

Proposition 6.9. *For all x in L, the linear map $\mathrm{ad}_x \colon L \to L$ is a derivation.*

This allows us to define a linear map $L \to \mathrm{Der}(L)$ by sending x to ad_x.

Proposition 6.10. *The map*

$$\mathrm{ad}\colon L \longrightarrow \mathrm{Der}(L),\ x \longmapsto \mathrm{ad}_x$$

is a morphism $L \to \mathrm{Der}(L)$ of Lie algebras.

Proof. The claim is equivalent to

$$\mathrm{ad}_{[x,y]} = [\mathrm{ad}_x, \mathrm{ad}_y], \tag{6.2}$$

which is easy to check: it is another form of the Jacobi identity. □

The image of L inside $\mathrm{Der}(L)$ consist of the *inner derivations* ad_x, the analogs of conjugation, and we can form the Lie algebras $\mathrm{Inn}(L)$ and $\mathrm{Out}(L)$ as for groups.

Exercise 30. Check that $\mathrm{Inn}(L)$ and $\mathrm{Out}(L)$ are well-defined Lie algebras. This means, in particular, the following: $\mathrm{Inn}(L) \leqslant \mathrm{Der}(L)$ is not only a subalgebra; it is an *ideal*. This is means that $[D, \mathrm{ad}_y]$ is inner for all derivations D: it equals $\mathrm{ad}_{D(y)}$, generalizing (6.2).

The situation is familiar from the theory of groups, where subgroups have to be normal if the quotient is to be a group again. We can also use this occasion to point out that there is no difference between left, right, and two-sided ideals in Lie algebras; there is only one notion, and it's called ideal.

Exercise 31. Check that the kernel of $\mathrm{ad}\colon L \to \mathrm{Der}(L)$ is the *center*

$$\mathrm{Z}(L) = \{x \in L \mid [x, y] = 0 \text{ for all } y \in L\}$$

of the Lie algebra L.

We therefore have an exact sequence

$$0 \longrightarrow \mathrm{Z}(L) \longrightarrow L \longrightarrow \mathrm{Der}(L) \longrightarrow \mathrm{Out}(L) \longrightarrow 0, \tag{6.3}$$

just as for groups.

6.5 Derivations on associative algebras

What does the exact sequence (6.3) look like when $L = A_{\mathrm{Lie}}$ is a Lie algebra underlying an associative algebra A?

Of course the term L becomes A_{Lie}, and the center of $L = A_{\mathrm{Lie}}$,

$$\begin{aligned} \mathrm{Z}(L) &= \{z \in A \,|\, [z, a] = 0 \text{ for all } a \in A\} \\ &= \{z \in A \,|\, za = az \text{ for all } a \in A\}, \end{aligned}$$

is the center of A in the usual sense.

We have to be more careful when it comes to derivations, though. It is straightforward to verify that here is an inclusion

$$\mathrm{Der}(A, \cdot) \xrightarrow{\subseteq} \mathrm{Der}(A_{\mathrm{Lie}}, [\,,\,]),$$

because $D(ab) = D(a)b + aD(b)$ implies $D[a, b] = [D(a), b] + [a, D(b)]$. However, there may be many more Lie derivations than associative derivations. For instance, this happens when A is commutative. This indicates that the smaller Lie algebras $\mathrm{Der}(A) = \mathrm{Der}(A, \cdot)$ is usually the more interesting object to study in the context of associative algebras.

6.6 Poisson algebras

A *Poisson algebra* is a Lie algebra that is also an associative algebra such that the bracket and the product are compatible: the bracket is a derivation, not just with respect to the bracket, but also with respect to the multiplication:

$$[x, y \cdot z] = [x, y] \cdot z + y \cdot [x, z]$$

Example 6.11. If A is an associative algebra, then A is automatically a Poisson algebra with respect to the commutator bracket.

Example 6.12. Let $A = \mathcal{C}^\infty(\mathbb{R})$ be the (commutative) algebra of smooth functions $\mathbb{R} \to \mathbb{R}$ with point-wise multiplications. We can identify a function f with the vector field $f\frac{\mathrm{d}}{\mathrm{d}x}$. The usual bracket of vector fields then gives a Lie bracket $\{?, ??\}$ on functions:

$$[f\frac{\mathrm{d}}{\mathrm{d}x}, g\frac{\mathrm{d}}{\mathrm{d}x}](h) = (fg' - f'g)\frac{\mathrm{d}}{\mathrm{d}x}h,$$

so that we get

$$\{f, g\} = fg' - f'g.$$

This satisfies

$$\{f, gh\} = \{f, g\}h + g\{f, h\},$$

so that $\{f, ?\}$ is a derivation for all f.

Example 6.13. The tensor algebra of a Lie algebra has a Poisson algebra structure.

6.7 A outlook on Hochschild cohomology

The center is the first (or rather, the zeroth) functor in a sequence of functors which form the Hochschild cohomology of the associative algebra A.

The Hochschild cohomology $\mathrm{HH}^\bullet(A; A)$ of an associative algebra A is the cohomology of the Hochschild cochain complex $\mathrm{CH}^\bullet = \mathrm{CH}^\bullet(A; A)$ with

$$\mathrm{CH}^n = \mathrm{Hom}(A^{\otimes n}, A),$$

and differential

$$\begin{aligned}\mathrm{d}(f)(a_1, \ldots, a_{n+1}) = {} & a_1 f(a_2, \ldots, a_{n+1}) \\ & + \sum_{j=1}^{n} (-1)^j f(a_1, \ldots, a_j a_{j+1}, \ldots, a_{n+1}) \\ & + (-1)^{n+1} f(a_1, \ldots, a_n) a_{n+1}.\end{aligned}$$

There are more conceptual definitions that explain the relevance of Hochschild cohomology, but the explicit version above has the advantage that we can immediately see

$$\mathrm{Z}(A) \cong \mathrm{HH}^0(A;A)$$

and

$$\mathrm{Out}(A) \cong \mathrm{HH}^1(A;A).$$

Exercise 32. For those who know about homology: verify these two isomorphisms.

The higher Hochschild cohomologies $\mathrm{HH}^n(A;A)$ for $n \geqslant 2$ are also important invariants of A, and they complete the picture. A good reference is Loday's book [Lod92], in particular its Section 1.5.

It turns out that the Hochschild cohomology of an associative algebra is both a commutative algebra and a Lie algebra in a graded sense: it is a Gerstenhaber algebra, a graded analogue of a Poisson algebra.

7 Universal enveloping algebras

It follows from Proposition 6.1 that there is a 'forgetful' functor from associative algebras to Lie algebras. It takes an associative algebra A with multiplication $(x, y) \mapsto x \cdot y$ to the the Lie algebra A with the commutator bracket

$$(x, y) \mapsto [x, y] = x \cdot y - y \cdot x.$$

In this section, we will find the best possible ('universal') functor, say U, in the other direction.

7.1 Universal enveloping algebras

We would like to associate an associative algebra $\mathrm{U}(L)$ with unit to a Lie algebra L in such a way that there is a natural (*adjunction*) bijection

$$\mathbf{Ass}(\mathrm{U}(L), A) \cong \mathbf{Lie}(L, A) \tag{7.1}$$

from the set of morphisms $\mathrm{U}(L) \to A$ to the set of morphism $L \to A$ that are compatible with the algebraic structures. The algebra $\mathrm{U}(L)$ is often called the *universal enveloping algebra* of L. We'll see several existence proofs below.

If we apply (7.1) to the algebra $A = \mathrm{U}(L)$, we see that the identity has to be mapped to morphism

$$L \to \mathrm{U}(L)_{\mathrm{Lie}}$$

of Lie algebras, where the target has the commutator bracket. This is the *unit* of the adjunction. One direction of the bijection (7.1) is given by restricting a morphism $\mathrm{U}(L)$ long the unit. Later, in Corollary 8.5, we shall come back to the question whether the unit is injective or not.

If we apply (7.1) to a Lie algebra of the form A_{Lie} with the commutator bracket, we see that the identity has to be mapped to a morphism

$$\mathrm{U}(A_{\mathrm{Lie}}) \to A$$

of associative algebras. This is the *co-unit* of the adjunction.

It is sometimes easy to calculate $\mathrm{U}(L)$ from L just using (7.1) and the fact that $\mathrm{U}(L)$ is determined by this equation.

Example 7.1. If $L = 0$, then there is a unique morphism $L \to A$ of Lie algebras for each associative algebra A so that there also has to be a unique morphism $\mathrm{U}(L) \to A$ for each of them, and this means

$$\mathrm{U}(0) \cong R,$$

the ground ring.

Example 7.2. For general reasons, an adjoint to a forgetful functor sends free objects to free objects: If $\mathrm{L}(V)$ is a free Lie algebra on a module V, then

$$\mathbf{Ass}(\mathrm{U}(\mathrm{L}(V)), A) \cong \mathbf{Lie}(\mathrm{L}(V), A) \cong \mathbf{Mod}(V, A) \cong \mathbf{Ass}(\mathrm{T}(V), A).$$

This shows that, for a free Lie algebra $\mathrm{L}(V)$, we have

$$\mathrm{U}(\mathrm{L}(V)) \cong \mathrm{T}(V),$$

the tensor algebra on V (see Appendix C).

For any Lie algebra L, we can form the tensor algebra $\mathrm{T}(L)$, and then there is a canonical embedding $L \to \mathrm{T}(L)_{\mathrm{Lie}}$. However, there is no reason why this should respect the Lie algebra structures. For that, we would need the Lie algebra bracket $[x, y]$ to be the same as the commutator bracket $x \otimes y - y \otimes x$ in the tensor algebra $\mathrm{T}(L)$, and they live in different degrees. However, if we force them to become equal, we obtain a construction of $\mathrm{U}(L)$:

Proposition 7.3. *For every Lie algebra L, there is an isomorphism*

$$\mathrm{U}(L) \cong \mathrm{T}(L)/([x, y] - x \otimes y + y \otimes x),$$

where the right hand side is the quotient algebra of the tensor algebra $\mathrm{T}(L)$ by the two-sided ideal generated by the elements $[x, y] - x \otimes y + y \otimes x$ for $x, y \in L$.

Proof. 'Compose' the universal properties of a tensor algebra and a quotient algebra. □

Example 7.4. If L is abelian, we get

$$\mathrm{U}(L) \cong \mathrm{T}(L)/(x \otimes y - y \otimes x) = \mathrm{S}(L),$$

the symmetric algebra on L. In particular, if L is abelian and free of rank n as a R–module, then $\mathrm{U}(L)$ is the polynomial algebra $R[x_1, \ldots, x_n]$ on n generators.

7.2 A glimpse into representation theory

One reason why it is useful to have enveloping algebras it is that they allow us to reduce certain aspects of Lie algebras to linear algebra in the sense of rings and modules. This applies, for instance, to representation theory: A *representation* of a Lie algebra L on a vector space V is a morphism $L \to \mathrm{gl}(V)$ of Lie algebras. Since $\mathrm{gl}(V) = \mathrm{End}(V)_{\mathrm{Lie}}$ is obtained from the algebra $\mathrm{End}(V)$ by forgetting structure, we see that it is essentially the same to have a morphism $\mathrm{U}(L) \to \mathrm{End}(V)$ of algebras, which is the same as a $\mathrm{U}(L)$–module structure on V. In summary, we can say that L–representations are the same thing as $\mathrm{U}(L)$–modules.

For the rest of this section, we will develop the representation theory of the Lie algebra sl_2 over an algebraically closed field of characteristic 0 through a sequence of exercises in linear algebra. This example plays an important role in the development of the general theory.

The Lie algebra sl_2 is 3–dimensional with basis

$$H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad X = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Exercise 33. Show that

$$[X, Y] = H, \quad [H, X] = 2X, \quad [H, Y] = -2Y.$$

An sl_2–representation is a vector space V with three operators H, X, Y that satisfy these three equations.

Let $v \in V$ be an eigenvector of H with eigenvalue λ. We often say that v has *weight* λ.

Exercise 34. Show that Xv is also an eigenvector, with weight $\lambda + 2$, and Yv is an eigenvector with weight $\lambda - 2$.

A non-zero vector v is called *primitive* if it is an eigenvector of H and if, in addition, we have $Xv = 0$. Note that we define primitive elements in bi-algebras later (Definition 10.1), and that is an unrelated notion.

Exercise 35. Show that every non-trivial, finite-dimensional representation has a primitive vector. Hint: start with an eigenvector of H and observe what happens when we apply X to it again and again.

Let v be a primitive vector of V of weight λ. Set $v_0 = v$, $v_1 = Yv$ and, in general,

$$v_j = \frac{Y^j}{j!} v.$$

Exercise 36. Compute the action of H, X, and Y on the v_j:

$$\begin{aligned} Hv_j &= (\lambda - 2j)v_j, \\ Xv_j &= (\lambda - j + 1)v_{j-1}, \\ Yv_j &= (j+1)v_{j+1}. \end{aligned}$$

It follows that the span of the v_j is a subrepresentation of V, and that

$$\{j \mid v_j \neq 0\} = \{0, 1, 2, \ldots, n\}$$

for some $n \geqslant 0$. Then $v_0, \ldots, v_n$ is a basis for this subrepresentation; it is of dimension $n + 1$, and its weight are

$$-n, -n + 2, \ldots, n - 2, n.$$

Let us write V_n for such a representation, unique up to isomorphism.

Example 7.5. The representation V_0 is the trivial 1–dimensional representation, with $H = X = Y = 0$.

Example 7.6. The representation V_1 is the 2–dimensional representation where each H, X, and Y acts as the $(2, 2)$–matrix we used to define it.

Example 7.7. The representation V_2 is the 3–dimensional representation of sl_2 on $V = \mathrm{sl}_2$ itself, where each H, X, and Y acts by the derivation that it defines. This is called the adjoint representation.

It turns out that all representations can be decomposed into the representations V_n in an essentially unique way.

8 The Poincaré–Birkhoff–Witt theorem

Let L be a Lie algebra over a ring K. The problem we face in this section is the following: how can we understand the universal enveloping algebra $\mathrm{U}(L)$ of L?

The universal enveloping algebra $\mathrm{U}(L)$ can be constructed as a quotient of the tensor algebra

$$\mathrm{T}(V) = \bigoplus_{j=0}^{\infty} \mathrm{T}^j(V).$$

We can use this to put a filtration on the universal enveloping algebra $\mathrm{U}(L)$ by setting

$$F^n\mathrm{U}(L)$$

to be the image of $\bigoplus_{j=0}^{n} \mathrm{T}^j(V)$ in $\mathrm{U}(L)$. This is an increasing filtration,

$$F^n\mathrm{U}(L) \leqslant F^{n+1}\mathrm{U}(L),$$

and it is exhausting:

$$\mathrm{U}(L) = \bigcup_n F^n\mathrm{U}(L).$$

The $F^n\mathrm{U}(L)$ are not ideals in $\mathrm{U}(L)$, but the filtration is multiplicative in the sense that

$$F^m\mathrm{U}(L) \cdot F^n\mathrm{U}(L) \leqslant F^{m+n}\mathrm{U}(L).$$

It follows that the associated graded algebra

$$\mathrm{GrU}(L) = \bigoplus_{n=0}^{\infty} \frac{F^n\mathrm{U}(L)}{F^{n-1}\mathrm{U}(L)}$$

is also an associated algebra with unit. The hope is that that this algebra is easier to describe than $\mathrm{U}(L)$, and that we can learn something about $\mathrm{U}(L)$ from $\mathrm{GrU}(L)$. These two wishes are fulfilled with the Poincaré–Birkhoff–Witt theorem.

8.1 Statement of the Poincaré–Birkhoff–Witt theorem

To state the Poincaré–Birkhoff–Witt theorem, it is helpful to make an easy observation.

Lemma 8.1. *For all Lie algebras L, the algebra $\mathrm{GrU}(L)$ is commutative.*

Proof. The tensor algebra $\mathrm{T}(L)$ is generated in degree 1 as an algebra. Threrefore, so are the universal enveloping algebra $\mathrm{U}(L)$ and its associated graded $\mathrm{GrU}(L)$. It follows that it is sufficient to show that the generators commute.

Given $x, y \in L$, their commutator

$$x \otimes y - y \otimes x = [x, y]$$

lies again in $F^1\mathrm{U}(L)$ and it is therefore zero in $F^2\mathrm{U}(L)/F^1\mathrm{U}(L) \leqslant \mathrm{GrU}(L)$. □

As a commutative algebra that is generated in degree 1 by L, the algebra $\mathrm{GrU}(L)$ admits a unique surjective (!) morphism

$$\mathrm{S}(L) \longrightarrow \mathrm{GrU}(L) \tag{8.1}$$

of algebras that is the identify on elements in degree 1 represented by elements of L.

Theorem 8.2. (Poincaré–Birkhoff–Witt) *If L is free as a K–module, then the morphism* (8.1) *is an isomorphisms, so that the associated graded of the universal enveloping algebra,*

$$\mathrm{GrU}(L) \cong \mathrm{S}(L),$$

is a polynomial algebra: the canonical morphisms (8.1) *is an isomorphism.*

Note that the hypothesis is automatically satisfied if K is a field, regardless of its characteristic.

Even though it might superficially seems that the Poincaré–Birkhoff–Witt theorem is more about $\mathrm{GrU}(L)$ than about $\mathrm{U}(L)$ itself, there is a lot that we can learn from it about the latter:

Corollary 8.3. *If L is free as a K–module with basis $x_1, \ldots, x_n$, then the universal enveloping algebra $\mathrm{U}(L)$ is free as a K–module with basis all monomials*

$$x_{i(1)} \cdot \cdots \cdot x_{i(r)}$$

for $1 \leqslant i(1) \leqslant \cdots \leqslant i(r) \leqslant n$.

Remark 8.4. There is a variant of the Poincaré–Birkhoff–Witt theorem that asserts the existence of a linear isomorphism $\mathrm{S}(L) \cong \mathrm{U}(L)$, without passage to the associated graded.

Corollary 8.5. *If L is free as a K–module, then the canonical morphisms*

$$L \longrightarrow \mathrm{U}(L)$$

is injective.

Example 8.6. If L is abelian, we have seen that $\mathrm{U}(L) \cong \mathrm{S}(L)$, and therefore the theorem is not a surprise.

Example 8.7. If $L = LA$ is free on A, we have seen that $\mathrm{U}(L) \cong \mathrm{T}(A)$, and therefore we get

$$\mathrm{Gr}\,\mathrm{T}(A) \cong \mathrm{S}(LA).$$

We will spend more time on this isomorphism later when discussing Witt results on free Lie algebras (see Section 11).

Remark 8.8. The hypothesis in Theorem 8.2 above is the one from the original papers of Birkhoff [Bir37] and Witt [Wit37]. Lazard [Laz54] and Cartier [Car58] have shown that the result also holds of K is a Dedekind domain, such as the ring of integers, without any hypothesis on L. Cohn [Coh63] has shown that the the conclusion holds if L is torsion-free as an abelian group, without any hypothesis on the ground ring K. Some hypothesis is necessary, as counterexamples due to

Cartier, Cohn, and Shirshov show. Higgins [Hig69] gave a unified exposition in which he replaced the Lie bracket and the universal enveloping algebra of L by a homological invariant, whose vanishing implies the Poincaré–Birkhoff–Witt theorem for L.

8.2 A proof of the Poincaré–Birkhoff–Witt theorem

The following proposition prepares the basic combinatorics of the argument that we will give.

Proposition 8.9. *Let L be a Lie algebra such that its underlying module is free with basis elements $x_1, x_2, \ldots$. Let V be the free module with basis elements v_J, where the $J = (j(1), \ldots, j(r))$ are finite sequences of increasing indices $j(1) \leqslant \cdots \leqslant j(r)$. Then there is representation of L on V such that*

$$x_i v_J = v_{(i,J)}$$

if $i \leqslant j(1)$ and, in general, $x_i v_J$ is a linear combination of v_K where K is at most one element longer than J.

Proof. Let us write $\ell(J) = r$ for the length of $J = (j(1), \ldots, j(r))$.

By (bi)linearity, we have to define $x_i v_J$ for all i and J. We can do that by induction on i and $\ell(J)$. We assume that it is already defined for i' and J' when $\ell(J') < \ell(J)$ or $\ell(J') = \ell(J)$ and $i' < i$.

He have to set $x_i v_J = v_{(i,J)}$ if $i \leqslant j(1)$ as required by the first condition in the statement.

If $i > j(1)$, then we write $J = (j(1), J')$ and set

$$x_i v_J = x_{j(1)} x_i v_{J'} + [x_i, x_{j(1)}] v_{J'}.$$

Note that $x_i v_{J'}$ is already defined by induction: $\ell(J') < \ell(J)$, and $x_i v_{J'}$ is a linear combination of v_k's with $\ell(K) \leqslant \ell(J)$. Note also that $j(1) < i$. We see that the

second condition in the statement is satisfied. This ends the definition of the $x_i v_J$ such that the conditions of the statement are satisfied.

It remains to be verified that we have a morphism $L \to \mathrm{gl}(V)$ of Lie algebras:

$$[x_i, x_j]v_K = x_i x_j v_K - x_j x_i v_K.$$

Both sides are alternating in x_i and x_j, so we can assume $i > j$.

If $j \leqslant k = k(1)$, then $x_j v_K = v_{(j,K)}$ and the equation above is the second case of the definition, with $j = j(1)$ and $J' = K$:

$$x_i x_j v_K = x_i v_{(j,K)} = x_j x_i v_K + [x_i, x_j]v_K.$$

We are left with $j > k$, so that $i > j > k$, and we have to check

$$x_i x_j x_k v_L - x_j x_i x_k v_L = [x_i, x_j]x_k v_L. \tag{8.2}$$

We will first do an induction on $\ell(L)$. Then, we already know

$$xyv_L = yxv_L + [x, y]v_L$$

for all x and y. The right hand side of the equation, for the elements $x = [x_i, x_j]$ and $y = x_k$, is

$$\begin{aligned}[x_i, x_j]x_k v_L &= x_k[x_i, x_j]v_L + [[x_i, x_j], x_k]v_L \\ &= x_k x_i x_j v_L - x_k x_j x_i v_L + [[x_i, x_j], x_k]v_L,\end{aligned}$$

whereas the left hand side is still

$$x_i x_j x_k v_L - x_j x_i x_k v_L.$$

Now we also want to do an induction on $\min\{i, j\}$. Since $k < \min\{i, j\}$, we know (8.2) for (j, k, i) and (k, i, j). Then equation (8.2) for (i, j, k) only is equivalent to the sum $(i, j, k) + (j, k, i) + (k, i, j)$ of the equations for (i, j, k), (j, k, i), and (k, i, j), cyclically permuting the indices. Adding things up, we see six terms on the left hand side, and the same six terms on the right hand side, plus the Jacobi identity, which is zero. □

Proposition 8.10. *Let V be the L–representation V constructed in Proposition 8.9. If we write $x_J = x_{j(1)} \cdots x_{j(r)}$ with $j(1) \leqslant \cdots \leqslant j(r)$, we have $x_J v_\varnothing = v_J$ for all sequences J.*

Proof. We can prove that by induction on the length $\ell(J)$ of J.

If the length is zero, then $J = \varnothing$ and $x_\varnothing = 1$.

If the length $\ell(J) \geqslant 1$, then we can write $J = (i, J')$ with $i \leqslant j'(1)$. Then we have $x_J = x_{j(1)} x_{J'}$. By induction, we get

$$x_J v_\varnothing = x_{j(1)} x_{J'} v_\varnothing = x_{j(1)} v_{J'},$$

and this is $v_{(j(1),J')} = v_J$ by the first property of the representation. □

Proof of Theorem 8.2. We'll actually prove the equivalent version given as Corollary 8.3: the elements

$$x_J = x_{j(1)} \cdots x_{j(r)}$$

with $j(1) \leqslant \cdots \leqslant j(r)$ form a basis of $\mathrm{U}(L)$. They clearly span $\mathrm{U}(L)$ since the x_j span L. To show their linear independence, we can use the morphism

$$\mathrm{U}(L) \longrightarrow \mathrm{End}(V)$$

of algebras induced by the representation V from Proposition 8.9, or rather the homomorphism

$$\mathrm{U}(L) \longrightarrow V,\ x_J \longmapsto x_J v_\varnothing$$

which is given by the action on the basis element $v_\varnothing$ for the empty sequence of length $\ell(\varnothing) = 0$. By Proposition 8.10, we have $x_J v_\varnothing = v_J$, and these are linearly independent by their definition. Then the same has to be true for the x_J. □

Remark 8.11. Bergman [Ber78] gives a proof of the Poincaré–Birkhoff–Witt theorem as an application of a more general result on normal forms in associative quotient algebras. (The Gröbner–Shirshov bases are an analog for commutative quotients algebras.) Bergman also mentions two open question for Lie algebras

over a field K: Does the existence of an isomorphism $\mathrm{U}(L) \cong \mathrm{U}(M)$ of associative algebras with unit imply the existence of an isomorphism $L \cong M$ of Lie algebras? Is a Lie algebra L free if the associative algebra $\mathrm{U}(L)$ with unit is free? This has been answered in the negative by Riley and Usefi [RU07], and in their example, one of the two Lie algebras is free. On the other hand, their base field is of positive characteristic. So the question might still be open in characteristic 0.

8.3 History

A bit of history might be in order for a theorem whose names are not in alphabetical order.

Borel [Bor01, p. 6] states that Poincaré, after some earlier work by Capelli on special cases, "more or less proved the Poincaré–Birkhoff–Witt theorem" in 1900. Ton-That–Tran [T-TT99] argue that Poincaré "gave a rigorous, complete, beautiful, and enlightening proof."

According to Schmid [Sch82], in 1937, Birkhoff [Bir37] and Witt [Wit37] "rediscovered Poincaré's result independently, in its most general version, needless to say–for possibly infinite-dimensional Lie algebras, over fields of arbitrary characteristic. Apparently it was Cartan–Eilenberg, in their book on homological algebra, who first affixed Poincaré's name to the theorem." The proof given in Cartan and Eilenberg [CE56, Theorem XIII.3.1] is modeled after Iwasawa [Iwa48], and so is the one in Bourbaki. Grivel [Gri04] compares the works of Poincaré, Birkhoff, and Witt, and gives a survey of later developments, in particular for Leibniz algebras (see Section 14.1).

9 Hopf algebras and groups

We have already seen very many different algebraic structures throughout the first part of these notes. What they all had in common was that they were defined in terms of sets X together with operations $X \times \cdots \times X \to X$ for various arities. Sometimes, as a modest variation, when we considered structures with an underlying linear structure and multilinear operations, we used the tensor product to express the operations in the form $X \otimes \cdots \otimes X \to X$. This was the case, for example, with associative algebras, commutative algebras, and Lie algebras.

In this section, we will review the *dual* situation, where the direction of the arrows is reversed and we have *co-operations* $X \to X \otimes \cdots \otimes X$ or even both kinds of operations. This leads to the Hopf algebra concept, which we can think of as a structure that unifies groups and Lie algebras.

The main result of the next section, Theorem 10.7, allows us to recover a Lie algebra from a Hopf algebra structure on its universal enveloping algebra, at least in good cases.

Some useful references for this section and the next are the expository papers by Bergman [Ber85], Serre [Ser93], and Cartier [Car07]. Kassel's textbook [Kas95] contains a lot of useful background material and discusses 'quantum groups:' a class of Hopf algebras that are similar to the ones that come from groups or Lie algebras. A good source for the history is the paper [AF09] by Andruskiewitsch–Ferrer Santos.

9.1 Co-algebras

Recall that we can describe an associative algebra A with unit as an R–module together with a multiplication

$$m\colon A \otimes A \longrightarrow A,\ a \otimes b \mapsto a \cdot b,$$

with $\otimes = \otimes_R$, and a unit

$$e\colon R \longrightarrow A,\ r \mapsto r \cdot 1$$

which are associative and unital in the sense that the diagram

$$\begin{array}{ccc} A \otimes A \otimes A & \xrightarrow{m \otimes A} & A \otimes A \\ {\scriptstyle A \otimes m}\downarrow & & \downarrow{\scriptstyle m} \\ A \otimes A & \xrightarrow[m]{} & A \end{array}$$

and the diagram

$$\begin{array}{ccccc} A & \xrightarrow{e \otimes A} & A \otimes A & \xleftarrow{A \otimes e} & A \\ & \searrow & \downarrow{\scriptstyle m} & \swarrow & \\ & & A & & \end{array}$$

commute.

Remark 9.1. Here and in the following we employ the Milnor–Moore convention: we write the symobol of an object A as short for the identity morphism id_A.

Remark 9.2. Note the we have suppressed the canonical identifications

$$A \otimes (A \otimes A) = A \otimes A \otimes A = (A \otimes A) \otimes A$$

and

$$R \otimes A = A = A \otimes R$$

from the diagrams.

We can easily dualize this definition:

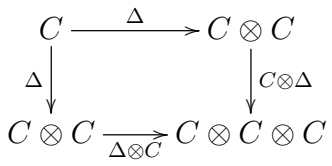
Definition 9.3. A *co-algebra* over a commutative ring R is an R–module C together with a *co-multiplication*

$$\Delta\colon C \longrightarrow C \otimes C$$

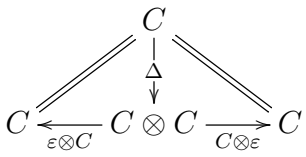
and a *co-unit*

$$\varepsilon\colon C \longrightarrow R$$

which are co-associative and co-unital in the sense that the diagram



and the diagram



commute.

Example 9.4. The ground ring R is a co-algebra in such a way that the structure morphisms are identities–or at least canonical identifications.

It is clear what a morphism $f : C \to D$ of co-algebras should be: a R–linear map that commutes with the structure maps.

Exercise 37. Show that the tensor product $C \otimes D$ of co-algebras C and D is again a co-algebra in a natural way.

Exercise 38. Dualize the definition of a module over an algebra A to arrive at a definition of a *co-module* over a co-algebra X.

We'll see more interesting examples of co-algebras soon, but first we'll combine the algebra and co-algebra concepts.

9.2 Bi-algebras

Definition 9.5. A *bi-algebra* B is an R–module together with R–linear maps

$$\begin{aligned} m\colon B \otimes B &\longrightarrow B \\ e\colon R &\longrightarrow B \\ \Delta\colon B &\longrightarrow B \otimes B \\ \varepsilon\colon B &\longrightarrow R \end{aligned}$$

such that (B, m, e) is an algebra and (B, Δ, ε) is a co-algebra, and furthermore these structures are compatible in the sense that Δ and ε are algebra morphisms.

Note that, when requiring that Δ and ε are algebra morphisms, we are using the usual algebra structure on R and on the tensor product $B \otimes B$.

Exercise 39. Show that m and e are morphisms of co-algebras. One way to do this is to write the requirements that Δ and ε are algebra morphisms as commutative diagrams and to interpret them in a different way.

9.3 Hopf algebras

Definition 9.6. A *Hopf algebra* is a bi-algebra $H = (H, m, e, \Delta, \varepsilon)$ together with an R–linear map

$$S\colon H \longrightarrow H,$$

the *antipode*, such that the diagram

$$\begin{array}{ccccc} H \otimes H & & \xrightarrow{S \otimes H} & & H \otimes H \\ {\scriptstyle\Delta}\uparrow & & & & \downarrow{\scriptstyle m} \\ H & \xrightarrow{\varepsilon} & R & \xrightarrow{e} & H \\ {\scriptstyle\Delta}\downarrow & & & & \uparrow{\scriptstyle m} \\ H \otimes H & & \xrightarrow[H \otimes S]{} & & H \otimes H \end{array}$$

commutes.

The antipode is not always a morphism of algebras or co-algebras.

9.4 Example: group algebras

Let X be a set. We denote by RX the free R–module with basis X. Its elements are the formal linear combinations

$$\sum_{x \in X} \lambda_x \cdot x$$

with $\lambda_x \in R$, almost all of them zero.

Remark 9.7. The functor $X \mapsto RX$ is left adjoint to the forgetful functor from the category of R–modules to the category of sets:

$$\mathrm{Hom}_R(RX, M) \cong \mathrm{Map}(X, M).$$

We have $R(\star) = R$ and

$$R(X \times Y) \cong RX \otimes RY,$$

more or less by the definition of the tensor product (see Appendix C). We can, therefore, use the unique map $X \to \star$ and the diagonal $\Delta \colon X \to X \times X$ to get two R–linear maps

$$\begin{aligned} \varepsilon &\colon RX \longrightarrow R\star = R,\ x \mapsto 1 \\ \Delta &\colon RX \longrightarrow R(X \times X) \cong RX \otimes RX,\ x \mapsto x \otimes x. \end{aligned}$$

These maps Δ and ε turn RX into a co-algebra over R. The co-unit ε sends a linear combination to the sum of its coefficients. The co-multiplication is co-commutative.

If $X = M$ is a monoid, the structure maps $m\colon M \times M \to M$ and $e\colon \star \to M$ induce two R–linear maps

$$\begin{aligned} e\colon R = R(\star) &\longrightarrow RM, \\ m\colon RM \otimes RM \cong R(M \times M) &\longrightarrow RM \end{aligned}$$

These maps m and e turn RM into an algebra over R. The multiplication is commutative if and only if M is commutative. Together with the co-algebra structure above, this turns RM into a bi-algebra.

If $M = G$ is a group the inversion $i\colon G \to G$ induces an R–linear map

$$S\colon RG \longrightarrow RG.$$

This turns RG into a Hopf algebra.

Example 9.8. The monoid algebra $R\mathbb{N}$ of the additive monoid of non-negative integers is the polynomial ring $R[T]$ in one variable T. We the group algebra $R\mathbb{Z}$ of the additive group of integers is the Laurent polynomial ring $R[T^{\pm 1}]$ in one variable T. If $G = \mathbb{Z}/n$ is cyclic of order n, then we get the quotient $R[T]/(T^n - 1)$.

Exercise 40. The functor $G \mapsto RG$ from the category of groups to the category of associative R–algebras with unit is left-adjoint to the functor from the category of associative R–algebras with unit to the category of groups that sends an algebra A to its group

$$A^\times = \{a \in A \,|\, \text{There is an } a' \text{ in } A \text{ such that } aa' = 1 = a'a.\}$$

of invertible elements.

Exercise 41. An element c in a coalgebra C is called *group-like* if $\varepsilon(c) = 1$ and $\Delta(c) = c \otimes c$. Show that the set $\mathrm{G}(C)$ of group-like elements in a co-algebra over a field K is K–linearly independent. This can be thought of as one variant of Dedekind's lemma on the linear independence of characters. To start, assume the contrary and choose a linear relation

$$c = \lambda_1 c_1 + \ldots \lambda_n c_n$$

between group-like elements. Inserting this into the equation $\Delta(c) = c \otimes c$, we get

$$\sum_j \lambda_i c_i \otimes c_i = \sum_{j,k} \lambda_j \lambda_k c_j \otimes c_k.$$

If the length of the relation was minimal, then the $c_j \otimes c_k$ are linearly independent in $C \otimes C$, and it follows that at most one of the λ_i can be non-trivial. Applying ε to $c = \lambda_i c_i$ shows $\lambda_i = 1$, which implies that $c = c_i$.

Exercise 42. In a group algebra KG, the elements of the basis G are group-like. It follows that these are the only ones: $G = \mathrm{G}(KG)$. Hence, one can recover a group from its group algebra as a Hopf algebra. Describe the group operation (Lemma 9.9). In the next section, we will see that we can similarly recover a Lie algebra from its universal enveloping algebra, at least in good cases.

Lemma 9.9. *In any bi-algebra, the product of two group-like elements is again a group-like element.*

Proof. Let g and h be group-like. Since the co-multiplication Δ is a morphism of algebras, we can compute

$$\Delta(g \cdot h) = \Delta(g) \cdot \Delta(h) = (g \otimes g) \cdot (h \otimes h) = g \cdot h \otimes g \cdot h.$$

The co-unit ε is also a morphism of algebras, and we get

$$\varepsilon(g \cdot h) = \varepsilon(g) \cdot \varepsilon(h) = 1 \cdot 1 = 1.$$

Both together prove the claim. □

9.5 Example: functions on groups

The class of examples that we study in this section will prepare us for a way to deal with groups in the later, more geometric parts of this course.

Once again, we start with a set X. We denote by $\mathcal{O}(X)$ the set of all functions $X \to R$ into our ground ring R. For example, the characteristic functions e_x

of the elements x of X are in $\mathcal{O}(X)$; these send x to 1 and all other elements to 0. Also, the constant functions are in $\mathcal{O}(X)$, too. Point-wise addition and multiplication turn $\mathcal{O}(X)$ into an R–algebra. This algebra is always commutative. As an R–module, it is dual to RX:

$$\mathcal{O}(X) = \mathrm{Map}(X, R) \cong \mathrm{Hom}_R(RX, R) = (RX)^\vee.$$

We have a canonical isomorphism $\mathcal{O}(\star) = R$, but the canonical embedding

$$\begin{aligned} \mathcal{O}(X) \otimes \mathcal{O}(Y) &\longrightarrow \mathcal{O}(X \times Y), \\ f \otimes g &\longmapsto (x, y) \mapsto f(x) \cdot g(y), \end{aligned}$$

which sends $e_x \otimes e_y$ to $e_{(x,y)}$, is *not* always an isomorphism. It is, however, if X and Y are finite. So we will assume for the rest of this section that the sets X we are considering are finite.

If $X = M$ is a finite monoid, the multiplication m and unit e induce two R–linear maps

$$\begin{aligned} \varepsilon\colon \mathcal{O}(M) &\longrightarrow \mathcal{O}(\star) = R, \\ \Delta\colon \mathcal{O}(M) &\longrightarrow \mathcal{O}(M \times M) \cong \mathcal{O}(M) \otimes \mathcal{O}(M) \end{aligned}$$

by precomposition. At the level of elements, these are given by

$$\begin{aligned} \varepsilon(f)(\star) &= f(1) \\ \Delta(f)(g, h) &= f(gh). \end{aligned}$$

Exercise 43. Show that

$$\Delta(e_g) = \sum_{ab=g} e_a \otimes e_b$$

by computing the values on all $(x, y) \in G \times G$.

The maps Δ and ε, together with m and e, turn $\mathcal{O}(M)$ into a bi-algebra over R.

If $M = G$ is a group, the inversion $i\colon G \to G$ induces an R–linear map

$$S\colon \mathcal{O}(G) \longrightarrow \mathcal{O}(G),$$

given by $(Sf)(g) = f(g^{-1})$. This turns $\mathcal{O}(G)$ into a Hopf algebra.

9.6 Polynomial functions on groups

Every commutative ring R has an underlying group, the additive group $\mathbb{G}_{\mathrm{a}}$; its ring of functions is defined to be the polynomial algebra

$$\mathcal{O}(\mathbb{G}_{\mathrm{a}}) = R[\mathrm{T}].$$

This algebra is a Hopf algebra with structure morphisms

$$\begin{aligned}\Delta(\mathrm{T}) &= \mathrm{T} \otimes 1 + 1 \otimes \mathrm{T} \\ \varepsilon(\mathrm{T}) &= 0 \\ S(\mathrm{T}) &= -\mathrm{T}\end{aligned}$$

Note that the canonical isomorphism

$$R[\mathrm{T}] \otimes R[\mathrm{T}] \cong R[\mathrm{U}, \mathrm{V}]$$

sends $\mathrm{T} \otimes 1 + 1 \otimes \mathrm{T}$ to $\mathrm{U} + \mathrm{V}$, making clear that the somewhat unusual expression $\mathrm{T} \otimes 1 + 1 \otimes \mathrm{T}$ models the universal formula for addition.

Every commutative ring R has a group of units, the multiplicataive group $\mathbb{G}_{\mathrm{m}}$; its ring of functions is defined to be the algebra of Laurent polynomials

$$\mathcal{O}(\mathbb{G}_{\mathrm{m}}) = R[\mathrm{T}^{\pm 1}].$$

This algebra is a Hopf algebra with structure morphisms

$$\begin{aligned}\Delta(\mathrm{T}) &= \mathrm{T} \otimes \mathrm{T} \\ \varepsilon(\mathrm{T}) &= 1 \\ S(\mathrm{T}) &= \mathrm{T}^{-1}\end{aligned}$$

Note that the canonical isomorphism

$$R[\mathrm{T}^{\pm 1}] \otimes R[\mathrm{T}^{\pm 1}] \cong R[\mathrm{U}^{\pm 1}, \mathrm{V}^{\pm 1}]$$

sends $\mathrm{T} \otimes \mathrm{T}$ to UV, making clear that this expression models the universal formula for multiplication.

10 Hopf algebras and Lie algebras

In this section we will prove Theorem 10.7, which allows us to recover a Lie algebra from the Hopf algebra structure on its universal enveloping algebra, at least in good cases.

10.1 Primitive elements

Definition 10.1. An element x of a bi-algebra B is called *primitive* if it satisfies the equation

$$\Delta(x) = x \otimes 1 + 1 \otimes x.$$

The set of primitive elements of B is denoted by $\mathrm{P}(B)$.

Remark 10.2. We could have added the condition $\varepsilon(x) = 0$ to the definition, but it is implied by the other equation. If $\Delta(x) = x \otimes 1 + 1 \otimes x$, then the co-unit axiom gives $x = x + e\varepsilon(x)$, so that $e\varepsilon(x) = 0$. We can then use that e is injective, because $\varepsilon e = \mathrm{id}$ by the bi-algebra axioms: the co-unit ε sends the unit e of B to the unit id of R.

Proposition 10.3. *The set $\mathrm{P}(B)$ of primitive elements of a bi-algebra B is a Lie algebra under the usual bracket $[x, y] = xy - yx$.*

Proof. It is obvious that $\mathrm{P}(B)$ is an R–submodule of B. It remains to be checked that it is closed under the bracket. This follows from expanding

$$\begin{aligned}\Delta[x, y] &= \Delta(x)\Delta(y) - \Delta(y)\Delta(x) \\ &= (x \otimes 1 + 1 \otimes x)(y \otimes 1 + 1 \otimes y) - \dots \\ &= xy \otimes 1 + x \otimes y + y \otimes x + 1 \otimes xy - \dots \\ &= [x, y] \otimes 1 + 1 \otimes [x, y].\end{aligned}$$

The terms $x \otimes y$ and $y \otimes x$ appear twice, with different signs. □

10.2 Universal enveloping algebras

Let L and M be Lie algebras. Then the product $L \times M$ is also a Lie algebra.

Exercise 44. Check that Lie algebra morphisms out of a product $L \times M$ can be described in a similar fashion as for groups (see Proposition B.1).

Exercise 45. Check there is a canonical isomorphism

$$\mathrm{U}(L \times M) \cong \mathrm{U}(L) \otimes \mathrm{U}(M),\ (x, y) \longmapsto x \otimes 1 + 1 \otimes y$$

of Lie algebras. The inverse is given by $a \otimes b \mapsto ab$. We can also verify that the right hand side satisfies the universal property of the left hand side.

The diagonal $\Delta_L \colon L \to L \times L$ is a morphism of Lie algebras. This induces a commutative diagram

$$\begin{array}{ccc} L & \xrightarrow{\Delta_L} & L \times L \\ \downarrow & & \downarrow \\ \mathrm{U}(L) & \xrightarrow[\mathrm{U}(\Delta_L)]{} & \mathrm{U}(L \times L), \end{array}$$

where the bottom arrow is an algebra morphism.

By composition, we obtain an algebra morphism

$$\Delta \colon \mathrm{U}(L) \longrightarrow \mathrm{U}(L \times L) = \mathrm{U}(L) \otimes \mathrm{U}(L).$$

There is also a canonical algebra homomorphism

$$\varepsilon \colon \mathrm{U}(L) \longrightarrow R,$$

the composition of the algebra morphism $\mathrm{U}(L) \to \mathrm{U}(0)$ induced by the Lie algebra morphism $L \to 0$, followed by the canonical identification $\mathrm{U}(0) = R$.

Proposition 10.4. *The algebra morphisms Δ and ε described above turn $\mathrm{U}(L)$ into a bialgebra. It is a Hopf algebra with antipode given by the morphism*

$$S \colon \mathrm{U}(L) \longrightarrow \mathrm{U}(L)$$

induced by the Lie algebra morphism $- \colon L \to L$.

Proof. It is easy to check equations between morphisms of algebras out of $\mathrm{U}(L)$: these are determined by their restriction to L. Note that Δ is a morphism of algebras *by construction*. □

Example 10.5. If L is abelian with basis $x_1, \dots$, then $\mathrm{U}(L) \cong \mathrm{S}(L)$ is the polynomial algebra $R[x_1, \dots]$, and the Hopf algebra structure is given as

$$\begin{aligned} \Delta(x_j) &= x_j \otimes 1 + 1 \otimes x_j \\ \varepsilon(x_j) &= 0 \\ S(x_j) &= -x_j \end{aligned}$$

We will later use that the Poincaré–Birkhoff–Witt isomorphism is compatible with the Hopf algebra structures in the following sense:

Proposition 10.6. *The Poincaré–Birkhoff–Witt homomorphism* $\mathrm{GrU}(L) \cong \mathrm{S}(L)$ *is compatible with the co-multiplications on both sides.*

Proof. There is nothing complicated going on here. Since Δ was constructed as an algebra morphism, it is determined by what it does to the generators in L. Then $\mathrm{U}(L) \to \mathrm{U}(L \times L)$ preserves the filtrations, and therefore passes on to a morphism between the associated graded algebras. These are polynomial, and Δ is still given by the same formula. □

Exercise 46. What about the co-unit and the antipode?

10.3 Primitive elements in universal enveloping algebras

By its construction, as in Exercise 45, we have $\Delta(x) = x \otimes 1 + 1 \otimes x$ for all x in L. This means that we have an inclusion $L \leqslant \mathrm{PU}(L)$ of L in the primitives of $\mathrm{U}(L)$. The following result states that, in most cases, more is true: we can recover a Lie algebra from its universal enveloping algebra as a bi-algebra: it is the Lie algebra of its primitives.

Theorem 10.7. *If L is a Lie algebra that is free as an R–module, for a ring R that itself is torsion-free, then we have an isomorphism*

$$L \cong \mathrm{PU}(L)$$

of Lie algebras.

Proof. We have already seen that there is an inclusion $L \leqslant \mathrm{PU}(L)$, and it remains to be seen that all primitives of $\mathrm{U}(L)$ lie in L.

Let us first deal with the case when L is abelian. Then $\mathrm{U}(L) \cong \mathrm{S}(L)$ is polynomial as an algebra, and the co-multiplication is given by substituting the sum: if we write the indeterminates as X_j, and $Y_j = X_j \otimes 1$ and $Z_j = 1 \otimes X_j$, then

$$\Delta(X_j) = X_j \otimes 1 + 1 \otimes X_j = Y_j + Z_j$$

for the generators. In general, we have

$$\Delta(f)(Y_j, Z_j) = f(Y_j + Z_j).$$

This means that f is primitive if and only if

$$f(Y_j + Z_j) = \Delta(f)(Y_j, Z_j) = (f \otimes 1 + 1 \otimes f)(Y_j, Z_j) = f(Y_j) + f(Z_j),$$

that is when f is additive. Since a polynomial is additive if and only if all of its homogeneous components are additive, we can assume that f is homogeneous of degree d. Then we get

$$2^d f(X_j) = f(2X_j) = f(X_j + X_j) = f(X_j) + f(X_j) = 2f(X_j).$$

By assumption on R, this means $f = 0$ or $2^d = 2$. But $2^d = 2$ can only be true for $d = 1$, that is when f lies in L, as claimed.

How does this help us if L is not abelian? In that case, the co-multiplication Δ on $\mathrm{U}(L)$ induces a co-multiplication $\mathrm{Gr}(\Delta)$ on $\mathrm{GrU}(L)$, and $\mathrm{GrU}(L) \cong \mathrm{S}(L)$ by the

Poincaré–Birkhoff–Witt theorem. So what we already have shown implies that the primitives of $\mathrm{U}(L)$ always lie in the first filtration $F^1\mathrm{U}(L)$.

Any element of the first filtration $F^1\mathrm{U}(L)$ can be written in the form $r + x$ with r in R and x in L. We have $\Delta(r) = r$ because Δ is a morphism of algebras, and for the co-multiplication on that we get

$$\Delta(r + x) = \Delta(r) + \Delta(x) = r + x \otimes 1 + 1 \otimes x.$$

if $r + x$ is primitive, then this equals

$$\begin{aligned}(r + x) \otimes 1 + 1 \otimes (r + x) &= r \otimes 1 + x \otimes 1 + 1 \otimes r + 1 \otimes x \\ &= 2r + x \otimes 1 + 1 \otimes x,\end{aligned}$$

and we must have $2r = r$, or $r = 0$. This shows that the primitive $r + x = x$ must lie in the Lie algebra L. □

11 Free Lie algebras

In this section, we will get a better understanding of the free Lie algebras. Here is a motivating (?) calculation.

Exercise 47. Let x and y be generators for the free Lie algebra on two generators. Then we have $[x, y]$ and $[y, x]$, but these two elements are linearly dependent because of $[y, x] = -[x, y]$. Similarly, we have eight 3–fold commutators $[x, [x, y]]$, $[y, [x, y]]$... involving only x and y, but they are all spanned by the first two. Check that! What do we get for 4–fold commutators? Can we find a linear relation between

$$[x, [x, [x, y]]], [x, [y, [x, y]]], [y, [x, [x, y]]], \text{ and } [y, [y, [x, y]]]?$$

We shall see that there must be one. (Hint: set $z = [x, y]$ in the Jacobi identity. This shows that the two terms in the middle are equal. Now that we know it: can we explain it without computation?) What do we get when we allow a third generator z?

We will find the dimension of the space of d–fold commutators in r free variables in this section.

11.1 Free algebras in general

First, let us look at constructions. Of course, free Lie algebras are defined by their universal property, but that does not mean that we cannot learn anything interesting from a construction of them. For instance, just as we have 'constructed' the free group on r generators in Section 2 as the 'set' of all r–ary operations on groups, we could 'construct' the free Lie algebra on r generators here as the 'set' of all r–ary operations on Lie algebras. Instead, we will take a different approach that works more generally. We start with some toy examples that do not involve an underlying linear structure.

Let X be a set. The free associative monoid with unit on X can be constructed as

$$\mathrm{Mon}(X) = \coprod_{d \geqslant 0} X^d = \star + X + (X \times X) + (X \times X \times X) + \cdots .$$

As a set, this $\mathrm{Mon}(X)$ is the sets of *words* on the alphabet X. Its elements are the sequences $(x_1, \ldots, x_d)$ of length d, with each x_j in X. Of course, if $d = 1$, then we simply write x, and not (x). The multiplication is given by concatenation

$$(x_1, \ldots, x_d)(y_1, \ldots, y_e) = (x_1, \ldots, x_d, y_1, \ldots, y_e).$$

The unique word of length $d = 0$, with no letter, is the unit.

Exercise 48. Show that a morphism $F\colon \mathrm{Mon}(X) \to N$ of associative monoids with unit is given uniquely by a map $f\colon X \to N$, via

$$F(x_1, \ldots, x_d) = (f(x_1), \ldots, f(x_d))$$

and $f = F|X$.

As a measure of complexity of $\mathrm{Mon}(X)$, let us count its elements. This fails in a naive sense because $\mathrm{Mon}(X)$ will be infinite if X is not empty. Instead, we can do the following. We note that the set $\mathrm{Mon}(X)$ is graded in the sense that there is a function

$$\ell\colon \mathrm{Mon}(X) \to \mathbb{N}$$

that sends a word $w = (x_1, \ldots, x_d)$ to its length $\ell(w) = d$. Let

$$\mathrm{Mon}(X)_d = \{w \in \mathrm{Mon}(X) \,|\, \ell(w) = d\}$$

be the subset of words of length d. If X is finite, then so is $\mathrm{Mon}(X)_d$, and we can consider the *Poincaré series*

$$P(t) = \sum_{d \geqslant 0} |\,\mathrm{Mon}(X)_d| \cdot t^d \in \mathbb{Z}[[t]],$$

a formal power series in a variable t. We'd like to find a nice expression for this power series as a count of the elements of $\mathrm{Mon}(X)$.

Write $|X| = r$ for the cardinality of X. Then $\mathrm{Mon}(X)_d = X^d$ has r^d elements, and consequently, we have

$$P(t) = 1 + rt + r^2t^2 + r^3t^3 + \cdots = \sum_{d \geqslant 0} r^d t^d = \frac{1}{1 - rt}. \tag{11.1}$$

That settles the case of the free associative monoid with unit.

The free commutative and associative monoid with unit on X can be constructed using the symmetric products

$$\mathrm{Sym}^d(X) = X^d/\Sigma_d$$

of the set X. Here, the symmetric group Σ_d acts on the words in X^d by permuting the letters. Then

$$\mathrm{Sym}(X) = \coprod_{d \geqslant 0} \mathrm{Sym}^d(X) \tag{11.2}$$

is a model for the free commutative and associative monoid with unit on X.

Exercise 49. Check this!

Remark 11.1. Note the similarity between (11.2) and the exponential series:

$$\mathrm{Sym}(X) = \star + X + \frac{X^2}{\Sigma_2} + \frac{X^3}{\Sigma_3} + \ldots,$$

and $|\Sigma_d| = d!$. In fact, there is also a natural bijection

$$\mathrm{Sym}(X + Y) \cong \mathrm{Sym}(X) \times \mathrm{Sym}(Y). \tag{11.3}$$

This is a consequence of the facts that (1) the functor Sym is a left-adjoint, so that it preserves coproducts, and (2) that $\times$ is the coproduct in the category of commutative monoids.

Despite what might have been suggested by the previous Remark 11.1, it is not true that the Poincaré series of $\mathrm{Sym}(X)$ is given by an exponential function. The reason is that the operation of the symmetric groups on the power sets are not free.

Lemma 11.2. *If X is a set with r elements, then*

$$|\operatorname{Sym}^d(X)| = \binom{r+d-1}{d}.$$

Proof. This is an easy counting argument. If $X = \{x_1, \dots, x_r\}$, then the elements of the symmetric power $\operatorname{Sym}^d(X)$ can be written, uniquely, as $[x_{j(1)}, \dots, x_{j(d)}]$, with

$$1 \leqslant j(1) \leqslant j(2) \leqslant \cdots \leqslant j(d) \leqslant r,$$

just by sorting the indices, using a permutation. We can 'stretch out' the indices to get

$$1 \leqslant j(1) < j(2) + 1 < \cdots < j(d) + d - 1 \leqslant r + d - 1,$$

so that the indices that correspond to a given element are given by a d–element subset of the set $\{1, \dots, r - d + 1\}$. □

As a consequence, we find the Poincaré series of $\operatorname{Sym}(X)$ to be

$$P(t) = \sum_{d \geqslant 0} \binom{r+d-1}{d} t^d$$

if $|X| = r$.

Exercise 50. Can we prove

$$\sum_{d \geqslant 0} \binom{r+d-1}{d} t^d = \left(\frac{1}{1-t}\right)^r$$

using (11.3)?

We will now turn to a more general structure, where we do not have associativity, as is the case for Lie algebras.

Definition 11.3. A *magma* is a set M that is equipped with a binary operation $m \colon M \times M \to M$. This operation does not have to satisfy any axiom.

Here is a construction of the free Magma $\mathrm{Mag}(X)$ on a set X of generators: We set

$$\mathrm{Mag}(X) = \coprod_n X_n,$$

where the sets X_n are defined for $n \geqslant 1$ inductively by $X_1 = X$ and

$$X_n = \coprod_{p+q=n} X_p \times X_q. \tag{11.4}$$

For instance, we have $X_2 = X \times X$,

$$X_3 = X \times (X \times X) + (X \times X) \times X,$$

and X_4 consists of the disjoint union of the five ways to bracket the four-fold power of X. The set $\mathrm{Mag}(X)$ becomes a magma with respect to the 'tautological' operation

$$m(x, y) = (x, y).$$

Remark 11.4. There is a relation with binary rooted trees with labels in X.

There is an obvious embedding $X \to \mathrm{Mag}(X)$, which identifies X with X_1. With respect to that map we have:

Proposition 11.5. *The magma* $\mathrm{Mag}(X)$ *is free with basis* X*.*

Proof. Every morphism $\mathrm{Mag}(X) \to N$ of magmas into a magma N restricts along the inclusion $X \to \mathrm{Mag}(X)$ to give a map $X \to N$. Conversely, given any map $f\colon X \to N$ into a magma N, there is a unique morphism $F\colon \mathrm{Mag}(X) \to N$ of magmas that extends it:

$$F(x, y) = (F(x), F(y)).$$

These two constructions are inverse to each other. □

Even if X has just one element, the set $\mathrm{Mag}(X)$ is infinite. Still, there is a way to measure the size of $\mathrm{Mag}(X)$ more precisely because it is graded:

Proposition 11.6. *If $|X| = r$, then*

$$|X_d| = \frac{1}{d}\binom{2d-2}{d-1}r^d$$

Equivalently, the Poincaré series is

$$P(t) = \sum_{d\geqslant 1} \frac{1}{d}\binom{2d-2}{d-1}r^d t^d.$$

The numbers occurring in this result are the *Catalan numbers*

$$c_n = \frac{1}{n+1}\binom{2n}{n}.$$

The Catalan numbers are ubiquitous in mathematics [Sta15]. They count the number of ways to put parentheses into a product of $n+1$ factors.

Example 11.7. We have

$$c_0 = 1, c_1 = 1, c_2 = 2, c_3 = 15, c_4 = 14...$$

Proof of Proposition 11.6. Since the factor r^d is rather obvious, it is enough to do the case $r = 1$. We need to find the Poincaré series

$$P(t) = t + t^2 + 2t^3 + 5t^4 + \dots$$

in that case. We see from (11.4) that this series satisfies

$$P(t) \cdot P(t) = t^2 + 2t^3 + 5t^4 + \cdots = P(t) - t,$$

or $P(t)^2 - P(t) + t = 0$. This means $P(t) = \frac{1}{2}(1 \pm \sqrt{1-4t})$. We can find the sign by setting $t = 0$: since there are no elements in $\mathrm{Mag}(X)$ of degree 0, we must have $P(0) = 0$. This gives

$$P(t) = \frac{1}{2}(1 - \sqrt{1-4t}).$$

The rest of the argument is calculus.

We start by using the binomial formula for the exponent $1/2$:

$$\begin{aligned}\sqrt{1-4t} &= \sum_{d\geqslant 0}\binom{1/2}{d}(-4t)^d \\ &= \sum_{d\geqslant 0}\frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\left(\frac{1}{2}-2\right)\cdots\left(\frac{1}{2}-d+1\right)}{d!}(-4)^d t^d \\ &= \sum_{d\geqslant 0}\frac{1\,(1-2)\,(1-4)\cdots(1-2d+2)}{d!}(-2)^d t^d \\ &= \sum_{d\geqslant 0}\frac{(-1)\cdot 1\cdot 3\cdot 5\cdots(2d-3)}{d!}2^d t^d.\end{aligned}$$

This gives

$$\frac{1}{2}(1-\sqrt{1-4t}) = \sum_{d\geqslant 1}\frac{1\cdot 3\cdot 5\cdots(2d-3)}{d!}2^{d-1}t^d.$$

Recall that we want to prove

$$P(t) = \sum_{d\geqslant 1}\frac{1}{d}\binom{2d-2}{d-1}t^d.$$

Therefore, it remains to verify

$$\frac{1\cdot 3\cdot 5\cdots(2d-3)}{d!}2^{d-1} = \frac{1}{d}\binom{2d-2}{d-1},$$

and we can do this as follows.

$$\begin{aligned}\frac{1}{d+1}\binom{2d}{d} &= \frac{(2d)!}{d!\cdot(d+1)!} \\ &= \frac{1\cdot 2\cdot 3\cdots(2d-1)\cdot 2d}{1\cdots d\cdot 1\cdots d\cdot(d+1)} \\ &= \frac{2^d\cdot 1\cdot 3\cdot 5\cdots(2d-1)}{1\cdots d\cdot(d+1)} \\ &= 2^d\frac{1\cdot 3\cdot 5\cdots(2d-1)}{(d+1)!}.\end{aligned}$$

□

11.2 Linearization

We have already seen a construction of the free associative R–algebra with unit on a set X: that is just the tensor algebra

$$\mathrm{T}(RX) = R\,\mathrm{Mon}(X),$$

and the Poincaré series is given as above.

We have also seen a construction of the free commutative (and associative) R–algebra with unit on a set X: that is just the symmetric algebra

$$\mathrm{S}(RX) = R\,\mathrm{Sym}(X),$$

and the Poincaré series is given as above.

It is formal that the free R–module $R\,\mathrm{Mag}(X)$ is the free R–algebra on the set X, where *R–algebra* just means R–module V with a R–linear multiplication $V \otimes V \to V$, no axioms required (not counting linearity).

11.3 Free Lie algebras

Now for the free Lie algebra, take the quotient of the free algebra by the alternating and the Jacobi relations:

$$\mathrm{L}(RX) = R\,\mathrm{Mag}(X)/I,$$

where I is the two-sided ideal generated by all elements of the form (x, x) and $(x, (y, z)) + (y, (z, x)) + (z, (x, y))$, where x, y, and z are elements in the set X. Of course, we will write $[x, y]$ for the Lie bracket on $\mathrm{L}(RX)$. The algebra $R\,\mathrm{Mag}(X)$ is graded, and the ideal I is generated by homogenous elements. Therefore, the free Lie algebra is also graded

$$\mathrm{L}(RX) = \bigoplus_{d \geqslant 1} \mathrm{L}^d(RX) = \mathrm{L}^1 X \oplus \mathrm{L}^2 X \oplus \cdots,$$

and each homogenous part $\mathrm{L}^d(RX)$ is finitely generated because $\mathrm{Mag}(X)_d$ is. Note that $\mathrm{L}^1X = RX$ and $\mathrm{L}^2X = \Lambda^2 RX$, but the higher L^dX for $d \geqslant 3$ are different from the usual multilinear functors.

We can now find the Poincaré series

$$P(t) = \sum_{d \geqslant 1} \operatorname{rank} \mathrm{L}^d(RX) t^d$$

of the free Lie algebra on a set X with r elements in the same way as we have found it for the free magma on a set X with r elements above. Of course, we have to assume that the 'rank' makes sense: for $R = \mathbb{Z}$ or R a field there is no problem. In the following, we'll suppress the ground ring R if confusion is unlikely.

11.4 The main result

Theorem 11.8. *The R–module L^dX is free of rank*

$$\frac{1}{d} \sum_{c|d} \mu(c) r^{d/c},$$

where μ is the Möbius function, so that the Poincaré series is

$$P(t) = \sum_{d \geqslant 1} \frac{1}{d} \sum_{c|d} \mu(c) r^{d/c} t^d.$$

Recall that the Möbius function is defined by $\mu(1) = 1$, $\mu(p_1 \dots p_k) = (-1)^k$ if the p_j are distinct primes, and $\mu(n) = 0$ otherwise.

I used the following GAP code to compute dimensions of the L^dX.

```
dimLie := function( r, d )
        D:= DivisorsInt( d );
        S:= Sum( D , c-> MoebiusMu( c ) * r^( d/c ) ) / d;
        return S;
end;;

L := List( [1..6], r-> List( [1..6], d -> dimLie( r, d ) ) );
Display( L );
```

Example 11.9. Let us look at small values for the number r of generators. If $r = 1$, the free Lie algebra on one generator x is just RX, with the trivial (abelian) bracket. GAP says

```
[    1,    0,    0,    0,    0,    0 ]
```

If $r = 2$, we have the free Lie algebra on two generators x and y, and GAP says:

```
[    2,    1,    2,    3,    6,    9 ]
```

confirming what we found in Exercise 47: there are is a 2–dimensional space of 3–fold commutators and a 3–dimensional space of 4–fold commutators in two generators. If $r = 3$, we have the free Lie algebra on three generator x, y, and z, and GAP says the following.

```
[    3,    3,    8,   18,   48,  116 ]
```

so that there should be 8–dimensional space of 3–fold commutators in three generators.

Example 11.10. On the other hand, we can look at small values of the degree d, and let the number of generators grow. If $d = 1$, we have

$$\operatorname{rank} \mathrm{L}^1 X = \frac{1}{1} \sum_1 \mu(1) r = r,$$

corresponding to $\mathrm{L}^1 X = RX$. If $d = 2$, we have

$$\operatorname{rank} \mathrm{L}^2 X = \frac{1}{2}\left(\sum_{c=1} \mu(c) r^2 + \sum_{c=2} \mu(c) r\right) = \frac{1}{2}(r^2 - r) = \binom{r}{2},$$

corresponding to $\mathrm{L}^2 X = \Lambda^2 RX$. If $d = 3$, we have

$$\operatorname{rank} \mathrm{L}^3 X = \ldots = \frac{1}{3}(r^3 - r).$$

For $r = 4$ we find three summands that need to be taken care of. In general, the value of $\operatorname{rank} \mathrm{L}^d X$ as a function of $|X| = r$ is a polynomial of degree d.

11.5 A proof of Theorem 11.8

Proof of Theorem 11.8. It suffices to show

$$d \cdot \operatorname{rank} \mathrm{L}^d X = \sum_{c|d} \mu(c) r^{d/c}, \tag{11.5}$$

In general, by Möbius inversion, an equation

$$f(d) = \sum_{c|d} \mu(c) g(d/c) = (\mu * g)(d)$$

is equivalent to

$$g(d) = \sum_{c|d} f(c) = (f * u)(d),$$

where u is the constant function 1; the unit with respect to $*$ is I, the characteristic function of 1, and μ is the inverse of u. It is, therefore, enough to prove

$$r^d = \sum_{c|d} c \cdot \operatorname{rank} \mathrm{L}^c X, \tag{11.6}$$

and that is what we will do.

We already know a lot. For instance, the Poincaré–Birkhoff–Witt theorem tells us that we have an isomorphism $\mathrm{GrU}(\mathrm{L}X) \cong \mathrm{S}(\mathrm{L}X)$, and we have seen in Example 7.2

that $\mathrm{U}(\mathrm{L}X) \cong \mathrm{T}(RX)$ formally. Also, the universal enveloping algebra $\mathrm{U}(\mathrm{L}X)$ and its associated graded $\mathrm{GrU}(\mathrm{L}X)$ have the same Poincaré series. Putting these facts together we see that

$$\mathrm{T}(RX) \text{ and } \mathrm{S}(\mathrm{L}X) \tag{11.7}$$

have the same Poincaré series. And we know them for both sides!

For the left side of (11.7), we get

$$\operatorname{rank} \mathrm{T}(RX) = \sum_{d=0}^{\infty} r^d t^d = \frac{1}{1-rt}.$$

As for the right side of (11.7), we can choose a basis $x_1, \dots$ of $\mathrm{L}X$. Note that this will be infinite in most cases. Also, the x_j will not just sit in degree 1. Let d_j be the degree of x_j. Then the Poincaré series of the symmetric algebra is

$$\operatorname{rank} \mathrm{S}(\mathrm{L}X) = \prod_j \frac{1}{1-t^{d_j}},$$

generalizing from Exercise 50 to the case when the variables not necessarily sit in degree 1. We collect the d_j for a given degree c to see that the Poincaré series is

$$\prod_{c=1}^{\infty} \frac{1}{(1-t^c)^{\operatorname{rank} \mathrm{L}^c X}}.$$

Equating the Poincaré series of both terms in (11.7) now gives

$$\frac{1}{1-rt} = \prod_{c=1}^{\infty} \frac{1}{(1-t^c)^{\operatorname{rank} \mathrm{L}^c X}}.$$

We apply $\log$ to this, remembering

$$\log\left(\frac{1}{1-y}\right) = \sum_{n=1}^{\infty} \frac{1}{n} y^n,$$

so that we get

$$\sum_{n=1}^{\infty} \frac{1}{n} r^n t^n = \sum_{c=1}^{\infty} \operatorname{rank} \mathrm{L}^c X \cdot \sum_{n=1}^{\infty} \frac{1}{n} t^{cn}.$$

Comparing the coefficients of t^d, we arrive at

$$\frac{1}{d} r^d = \sum_{\substack{c,n \\ cn=d}} \operatorname{rank} \mathrm{L}^c X \cdot \frac{1}{n},$$

or

$$r^d = \sum_{c|d} \operatorname{rank} \mathrm{L}^c X \cdot c,$$

which is precisely what we wanted. □

12 Free groups revisited

In this section, we determine the associated graded Lie algebras of free groups: these are free Lie algebras. In contrast to the statement that the universal enveloping algebra of a free Lie algebra is free, which we saw in Example 7.2 follows formally from adjunctions, the result of the present section is not formal, and requires a more involved proof.

Let $\mathrm{F}X$ denote a free group on the set X. The aim of this section is to describe the associated graded

$$\operatorname{Gr}\mathrm{F}X = \bigoplus_{n=1}^{\infty} \Gamma_n(\mathrm{F}X)/\Gamma_{n+1}(\mathrm{F}X)$$

as a Lie algebra. The free group comes with a canonical map $X \to \mathrm{F}X$ that sends the elements of X to the generators of $\mathrm{F}X$, and this induces a map $X \to \operatorname{Gr}\mathrm{F}X$. As the target is a Lie algebra, we get a canonical morphism $\mathrm{L}X \to \operatorname{Gr}\mathrm{F}X$ of Lie algebras. This morphism is surjective, because the target is generated, as a Lie algebra, by the elements of X.

Theorem 12.1. *The canonical surjection*

$$\mathrm{L}X \xrightarrow{\cong} \operatorname{Gr}\mathrm{F}X$$

is an isomorphism of Lie algebras.

This isomorphism respects the gradings on both sides. Therefore, from the results of the previous section, we get:

Corollary 12.2. *The free abelian group* $\Gamma_n(\mathrm{F}X)/\Gamma_{n+1}(\mathrm{F}X)$ *has rank*

$$\frac{1}{n}\sum_{d|n} \mu(d) r^{n/d}$$

when $|X| = r$.

Before we begin with a proof, we introduce a useful device.

12.1 A completion

We look at the tensor algebra

$$\mathrm{T}\mathbb{Z}X = \bigoplus_{n=0}^{\infty} (\mathbb{Z}X)^{\otimes n},$$

the free associative R–algebra with unit on X. We replace the sum by the product,

$$\widehat{\mathrm{T}}\mathbb{Z}X = \prod_{n=0}^{\infty} (\mathbb{Z}X)^{\otimes n},$$

and refer to this as the *completion* $\widehat{\mathrm{T}}\mathbb{Z}X$ of $\mathrm{T}\mathbb{Z}X$. Its elements $f = (f_n \,|\, n \geqslant 0)$ are the infinite sequences with $f_n \in (\mathbb{Z}X)^{\otimes n}$. The addition is termwise, and the multiplication is given by

$$(f \cdot g)_n = \sum_{j+k=n} f_j g_k.$$

We have an inclusion $\mathrm{T}\mathbb{Z}X \leqslant \widehat{\mathrm{T}}\mathbb{Z}X$, and the image consists of those sequences that are eventually zero.

Example 12.3. If $x \in X$ is a generator, then $1 + x \in \mathrm{T}\mathbb{Z}X$. This element is invertible in the completion $\widehat{\mathrm{T}}\mathbb{Z}X$: the inverse is given by

$$((-x)^n \,|\, n \geqslant 0) = 1 - x + x^2 - x^3 + \ldots,$$

and this inverse lies in the complement $\widehat{\mathrm{T}}\mathbb{Z}X \setminus \mathrm{T}\mathbb{Z}X$, unless x is nilpotent in the tensor algebra, which only happens if x is zero.

The preceeding example shows that we have a map $X \to \left(\widehat{\mathrm{T}}\mathbb{Z}X\right)^{\times}$ in the group of units of the completion. It follows from the universal property of the free group $\mathrm{F}X$ that we get a morphism

$$\theta \colon \mathrm{F}X \longrightarrow \left(\widehat{\mathrm{T}}\mathbb{Z}X\right)^{\times}$$

of groups.

There is an obvious filtration

$$M^n = \{f \in \widehat{\mathrm{T}}\mathbb{Z}X \mid f_0 = \cdots = f_{n-1} = 0\}$$

on the completion, and we use that to define a 'new' filtration

$$U^n = \theta^{-1}(1 + M^n) \leqslant \mathrm{F}X$$

on the free group. An element $g \in \mathrm{F}X$ lies in U^n if and of if $\theta(g)$ has the form

$$\theta(g) = 1 + f_n + f_{n+1} + \dots.$$

Remark 12.4. We will later prove, in Proposition 12.7, that this 'new' filtration agrees with the lower central series filtration Γ_n, but we do not know this yet. We will have to prove some of the properties for U that we already know for Γ. For instance, we obviously have

$$\mathrm{F}X = \mathrm{U}^1 \geqslant U^2 \geqslant \cdots \geqslant \bigcap_{n=1}^{\infty} U^n = \{e\}.$$

Here, we know that we have $U^1 = \mathrm{F}X$ because U^1 contains the generators.

Lemma 12.5. *We have* $[U^m, U^n] \leqslant U^{m+n}$.

Proof. For $g \in U^m$ we have $\theta(g) = 1 + a$ with $a \in M^m$, and similarly for $h \in U^n$ we have $\theta(h) = 1 + b$ with $b \in M^n$. As θ is a morphism of groups, we get

$$\theta(gh) = 1 + a + b + ab$$
$$\theta(hg) = 1 + a + b + ba,$$

and we would like to use this to calculate $\theta[g, h]$. Recalling

$$[g, h] = ghg^{-1}h^{-1} = (gh)(hg)^{-1},$$

we have to calculate $\theta(gh)\theta(hg)^{-1}$. Now $\theta(gh) = 1 + s$ with $s = a + b + ab$ and similarly $\theta(gh) = 1 + t$ with $t = a + b + ba$. We get

$$\begin{aligned}
\theta[g, h] &= (1 + s)(1 + t)^{-1} \\
&= (1 + s)(1 - t + t^2 - t^3 + \dots) \\
&= 1 + (s - t) - (s - t)t + (s - t)t^2 - \dots \\
&= 1 + (s - t) + \dots \text{ (terms of higher degree).}
\end{aligned}$$

This means that

$$\theta[g,h] = 1 + (ab - ba) + \dots \tag{12.1}$$

with $ab - ba \in M^{m+n}$, so that $[g,h] \in U^{m+n}$. □

It follows from this lemma and Proposition 4.7, the minimality of the lower central series, that we have

$$\Gamma_n(\mathrm{F}X) \leqslant U^n$$

for all $n \geqslant 1$.

Let us write

$$\mathrm{Gr}_U \mathrm{F}X = \bigoplus_{n=1}^{\infty} U^n/U^{n+1}$$

for the associated graded Lie algebra with respect to the filtration U. For clarity, we will also write $\mathrm{Gr}_\Gamma \mathrm{F}X = \mathrm{Gr}\,\mathrm{F}X$ for the Lie algebra that we are originally interested in.

Lemma 12.6. *There is a morphism*

$$\mathrm{Gr}_U \mathrm{F}X \longrightarrow \mathrm{T}\mathbb{Z}X$$

of Lie algebras sending the class $[g]$ of an element $g \in U^n$ to the element a_n that is given by the equation

$$\theta(g) = 1 + a_n + a_{n+1} + \dots .$$

Proof. It is obvious that this is well-defined, and (12.1) shows that it is a morphism of Lie algebras. □

Proof of Theorem 12.1. We can consider the diagram

$$\begin{array}{ccc} \mathrm{Gr}_\Gamma \mathrm{F}X & \longrightarrow & \mathrm{Gr}_U \mathrm{F}X \\ \uparrow & & \downarrow \text{Lemma 12.6} \\ \mathrm{L}X & \dashrightarrow & \mathrm{T}\mathbb{Z}X \end{array}$$

of morphisms of Lie algebras.

The top arrow is induced by the minimality of the lower central series. The left arrow is the one that we want to prove is bijective; we already know that it is surjective, so that it remains to show injectivity. The bottom arrow is defined as the composition of the other three arrows, so that the diagram automatically commutes.

There is also another description of the bottom arrow, using the universal property of the free Lie algebra $\mathrm{L}X$: it has to be the canonical morphism $\mathrm{L}X \to \mathrm{T}\mathbb{Z}X$ coming from the adjunction, as we can check on generators. We have already seen in Corollary 8.5 that this morphism is injective. It follows that also the first map $\mathrm{L}X \to \mathrm{Gr}_\Gamma \mathrm{F}X$ in the composition must be injective, as desired. □

We now address the problem left open in Remark 12.4

Proposition 12.7. *The filtration U agrees with the lower central series filtration.*

Proof. We show $U^n = \Gamma_n(\mathrm{F}X)$ by induction on n. For $n = 1$ this can be verified directly, because $U^1 = \mathrm{F}X = \Gamma_1(\mathrm{F}X)$. Let us, therefore, assume that $n \geqslant 2$. From the proof of Thereom 12.1, we already know that

$$\Gamma_{n-1}(\mathrm{F}X)/\Gamma_n(\mathrm{F}X) \longrightarrow U^{n-1}/U^n$$

is injective. By induction we have $\Gamma_{n-1}(\mathrm{F}X) = U^{n-1}$. It follows that this map is also surjective. The claim now follows from the 5–Lemma. □

13 Exponentials

In this section, we will think about the exponential series

$$\exp(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!},$$

and, therefore, we will have to assume that we are working over some commutative ring with unit that contains the rationals $\mathbb{Q}$. The inverse is the logarithm

$$\log(1+t) = \sum_{l=1}^{\infty} \frac{(-1)^{l+1}}{l} t^l,$$

because $\exp(\log(1+t)) = 1+t$ and $\log(\exp(t)) = t$ in $\mathbb{Q}[[t]]$.

Exercise 51. Show that

$$\exp(s)\exp(t) = \exp(s+t)$$

whenever s and t commute, such as $s, t \in \mathbb{Q}[[s,t]]$, the power series ring.

Exercise 52. Find two $(2,2)$–matrices S and T such that

$$\exp(S)\exp(T) \neq \exp(S+T).$$

Hint: by the previous exercise, we have to find a pair of matrices that does *not* commute. For instance, we can try

$$S = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Then

$$ST = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{and} \quad TS = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

show that these do not commute. For the exponentials, we find

$$\exp(S) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \exp(T) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

so that

$$\exp(S)\exp(T) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix},$$

while

$$\exp(S+T) = \begin{pmatrix} \cosh(1) & \sinh(1) \\ \sinh(1) & \cosh(1) \end{pmatrix}.$$

When $\exp(x)\exp(y) \neq \exp(x+y)$, the question is: are there other ways of expressing the result? In other words, we would like to set

$$F(x,y) = \log(\exp(x)\exp(y))$$

and find a 'nice' formula for it. First of all, we'll have to make sense of the formula at all: where does it live?

We need powers of x and y, and at least a formal way of convergence for that. Let us assume that we ignore power of order 4 or higher, so that

$$\exp(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3$$

and similarly for $\exp(y)$. Paying attention to the order of things, we then get

$$\exp(x)\exp(y) = ... = 1 + (x+y) + \frac{1}{2}(x^2 + 2xy + y^2) + \frac{1}{6}(x^3 + 3x^2y + 3xy^2 + y^3),$$

where we can't simplify $x^2+2xy+y^2$ to $(x+y)^2$, unless x and y commute! Applying

$$\log(1+z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3,$$

we arrive at

$$F(x,y) = ... = x + y + \frac{1}{2}[x,y] + \frac{1}{12}\left([x,[x,y]] + [y,[y,x]]\right).$$

This suggests that we can, in general, express $F(x,y)$ in terms of the free Lie algebra generated by x and y. The main theorem below will say that this is indeed the case.

Since we do care about arbitrary high powers, we can once again use the completion

$$\widehat{\mathrm{T}}(X) = \prod_{n=0}^{\infty} \mathrm{T}^n(X),$$

where X is a set of variables, such as $\{x, y\}$, for instance. The algebra $\widehat{\mathrm{T}}(X)$ contains

$$\widehat{\mathrm{L}}(X) = \prod_{n=0}^{\infty} \mathrm{L}^n(X),$$

the *completion* of the free Lie algebra on X, and it can be identified with the primitive elements in the Hopf algebra $\widehat{\mathrm{T}}(X)$, the completion of the universal enveloping algebra of $\widehat{\mathrm{L}}(X)$.

We let $\widehat{M}$ denote the ideal

$$\widehat{M} = \prod_{n=1}^{\infty} \mathrm{T}^n(X)$$

of $\widehat{\mathrm{T}}(X)$ generated by the elements X in degree 1. There are obvious bijections

$$\widehat{M} \longleftrightarrow 1 + \widehat{M}$$

given by adding and subtracting 1, but we can also use the exponential and the logarithm:

Proposition 13.1. *The exponential and the logarithm define bijections*

$$\exp\colon \widehat{M} \longleftrightarrow 1 + \widehat{M} \colon \log.$$

Proof. It is easy to see that these define maps in the indicated directions. The fact that these are inverse to each other follows from the power series identities in $\mathbb{Q}[[t]]$. □

Proposition 13.2. *The bijections in Proposition 13.1 induce bijections between the primitive and group-like elements:*

$$\exp\colon \{p \in \widehat{M} \mid \mu(p) = p \otimes 1 + 1 \otimes p\} \longleftrightarrow \{g \in 1 + \widehat{M} \mid \mu(g) = g \otimes g\} \colon \log$$

Proof. Let $g = \exp(p)$ with p primitive. We have to show that g is group-like. Since the co-multiplication is an algebra morphisms, it commutes with the exponential. This gives

$$\mu(g) = \mu(\exp(p)) = \exp(\mu(p)) = \exp(p \otimes 1 + 1 \otimes p).$$

The elements $p \otimes 1$ and $1 \otimes p$ commute, because their product is $p \otimes p$ in either order. Then we can continue

$$\begin{aligned}\exp(p \otimes 1 + 1 \otimes p) &= \exp(p \otimes 1) \cdot \exp(1 \otimes p) \\ &= (\exp(p) \otimes 1) \cdot (1 \otimes \exp(p)) \\ &= (g \otimes 1) \cdot (1 \otimes g) \\ &= g \otimes g,\end{aligned}$$

as claimed.

The argument for $p = \log(g)$ is similar. □

Theorem 13.3. (Baker–Campbell–Hausdorff) *The element*

$$\log(\exp(x)\exp(y)) \in \widehat{\mathrm{T}}\{x, y\}$$

lies in $\widehat{\mathrm{L}}\{x, y\}$.

Proof. It is enough to show that $\log(\exp(x)\exp(y))$ is primitive. By the previous proposition, we can also show that $\exp(x)\exp(y)$ is group-like. We can use, again, that μ is a morphism of algebras, so that μ commutes with $\exp$. Also, the elements x and y are primitive because they lie in the Lie algebra, so that $\exp(x)$ and $\exp(y)$ are group-like. This gives

$$\begin{aligned}\mu(\exp(x)\exp(y)) &= \mu(\exp(x))\mu(\exp(y)) \\ &= \mu(\exp(x))\mu(\exp(y)) \\ &= (\exp(x) \otimes \exp(x)) \cdot (\exp(y) \otimes \exp(y)) \\ &= \exp(x)\exp(y) \otimes \exp(x)\exp(y),\end{aligned}$$

as claimed. Of course, this is a variant of the more abstract observation that a product of group-like elements is group-like (Lemma 9.9). □

Remark 13.4. As perhaps already suggested by the length of its name, the Baker–Campbell–Hausdorff theorem has a long history, with contributions also by Poincaré and Schur. We refer to [AB12, BF12] for this, and to [Eic68, Tu85, Tu04] for some more recent proofs. We have only seen the result in its qualitative form, but there are versions with explicit formulas for $F(x, y)$, the first one being due to Dynkin.

14 Fade out: related algebraic structures

In this section, we will briefly introduce a few more algebraic structures that are related to Lie algebras and that help us understand various aspects of the Jacobi identity.

14.1 Leibniz and Zinbiel algebras

A *Leibniz algebra* is an algebra $(A, \star)$ such that all left-multiplications are derivations:

$$x \star (y \star z) = (x \star y) \star z + y \star (x \star z). \tag{14.1}$$

For instance, every Lie algebra defines a Leibniz algebra via $x * y = [x, y]$. But note that $x \star x = 0$ is not required for a Leibniz algebra. Therefore, one might say that Leibniz algebras are like "non-commutative Lie algebras."

More precisely, the equation (14.1) defines *left* Leibniz algebras. *Right* Leibniz are defined using the equation

$$(x \star y) \star z = (x \star z) \star y + x \star (y \star z),$$

which says that all right-multiplications are derivations. Both equations say that the associator can be written as a triple-product.

Loday's papers [Lod93, Lod03] indicate why Leibniz algebras are a fundamental algebraic structure. What he does not say is that Leibniz algebras can be thought of as a Lie analogs of racks (as in Exercise 4).

The equation

$$(x \star y) \star z = x \star (y \star z) + x \star (z \star y)$$

defines *Zinbiel* algebras. 'Zinbiel' is 'Leibniz' read backward.

Exercise 53. If $(A, \star)$ is a Zinbiel algebra, then

$$ab = a \star b + b \star a$$

is an associative product on A. It is also commutative, rather obviously.

Part II: Differential Equations

15 Differentials and differential equations

We would like to study the symmetry groups of differential equations. We need a bit of terminology to make this precise. Before we come to that, it should be useful to point out the following analogy: the classical Galois theory studies the symmetry groups of polynomial equations; and the elements of Galois groups permute the solutions to the equations. For differential equations, the symmetry groups will act on the solution spaces via linear transformations. What both theories share is the insight that it is helpful to consider the fields generated by the numbers or functions that solve the equations.

15.1 Differential fields

The following definition presents a very natural mix of the notion of a field and that of a derivation.

Definition 15.1. A *differential field* is a field F together with a derivation: an additive map $D\colon F \to F$ such that $D(fg) = D(f)g + fD(g)$ holds for all f, g in F.

Note that is does not make sense to require that D is F–linear because there are too few interesting F–linear maps $F \to F$.

Exercise 54. Find all F–linear derivations $F \to F$.

Exercise 55. Let $F[\varepsilon]/(\varepsilon^2)$ be the ring of dual numbers over the field F. Then there is a bijection between the set of derivations on F and the set of ring (or F–algebra) morphisms $F \to F[\varepsilon]/(\varepsilon^2)$ over F: the dotted arrows that make the diagram

$$\begin{array}{ccc} F & \dashrightarrow & F[\varepsilon]/(\varepsilon^2) \\ & \searrow & \downarrow{\scriptstyle a_0+a_1\varepsilon\mapsto a_0} \\ & & F \end{array}$$

commute. It is given by sending a derivation D to the map $a \mapsto a + D(a)\varepsilon$.

Exercise 56. Show that

$$\{a \in F \mid D(a) = 0\}$$

is a subfield K of F. This is the *field K of constants* C_F with respect to D. Show that a derivation D is linear for its field of constants.

Let's start with a silly example:

Example 15.2. Every field F becomes a differential field with respect to the zero derivation $D = 0$. In that case, the field of constants is just $\mathrm{C}_F = F$ itself. Of course, the zero-derivation is not very interesting. The example only shows that any field admits a differential and shows up as a field of constants.

Here is a short list of the basic differential fields that we will be working with.

Example 15.3. The field $\mathbb{C}(t)$ of rational functions is the fraction field of the ring $\mathbb{C}[t]$ of polynomials. Its elements are the fractions

$$\frac{p(t)}{q(t)}$$

of polynomials, with the usual addition and multiplication, and the usual derivation. This pair gives a differential field. Its field of constants is $\mathbb{C}$.

Example 15.4. Let $U \subseteq \mathbb{C}$ be a non-empty open subset. Let $\mathcal{O}(U)$ denote the ring of holomorphic functions on U. If U is convex, then U has no zero-divisors, and we can pass to the fraction field $\mathcal{M}(U)$ of meromorphic functions on U. Convexity is only convenient; we could get away with less strict assumptions, but our goal is not maximum generality, as long as we can deal with the main examples. Fields like $\mathcal{M}(U)$ are also examples of differential fields, and they contain the differential fields $\mathbb{C}(t)$.

Example 15.5. Every holomorphic function can be expanded in a power series around 0 say. The fraction field $\mathbb{C}((t))$ of the formal power series ring $\mathbb{C}[[t]]$ is the ring of formal Laurent series.

15.2 Linear differential equations

Given a differential field (F, D), a *linear differential operator of degree* r is a polynomial in D with coefficients in F: a map $L\colon F \to F$ of the form

$$L(u) = f_r D^r(u) + \cdots + f_1 D(u) + f_0 u \tag{15.1}$$

with $f_j \in F$ and $f_r \neq 0$. Of course, since we are working over a field, we can divide by f_r and assume that $f_r = 1$. If all coefficients f_j lie in the field K of constants, then we say that L has *constant coefficients*.

Example 15.6. The derivation D itself is a first order linear differential operator. Its solutions define the field of constants of F.

If U convex, or more generally simply-connected, then the field $\mathcal{M}(U)$ is large enough to contain solutions to all such linear differential equations of finite order:

Theorem 15.7. *For any given point z_0 in U, and any given initial condition $D^j u(z_0)$ for $j = 0, \ldots, r-1$, there exists a unique solution of the equation* (15.1)*.*

A proof is contained in [BR89], for instance.

The solution space to linear differential operator L of degree r is a vector space V_L over the field $K = \mathbb{C}$ of constants of dimension r. The initial conditions at z_0 give an isomorphism $V_L \cong \mathbb{C}^r$, but such an isomorphism is not canonical: it depends on the point z_0.

Definition 15.8. The smallest differential subfield of $\mathcal{M}(U)$ that contains all solutions the the equation (15.1) is called the *differential splitting field* E_L of the linear differential operator L.

If u is a solution of the r–th order linear differential equation $L(u) = 0$, then the splitting field contains $u, D(u), D^2(u), \ldots, D^{r-1}(u)$ and all rational functions in these. If $u_1, \ldots, u_r$ is a basis of the solution space V_L, then the r^2 functions

$$D^j(u_k)$$

for $0 \leqslant j \leqslant r-1$ and $1 \leqslant k \leqslant r$ generate the splitting field E_L as a field.

16 Quadratures

In this section, we will describe the "elementary" methods of solving differential equations: by integration, by exponentiation, and by algebraic methods.

16.1 Integrals

Let f be a function in f that is not a derivative; it is not an integral. There is an extension E that contains an integral u, and we can assume that E is generated by u as a differential field. The integral u solves the equation

$$u' = f.$$

This equation does *not* have the form (15.1). It is an inhomogenous equation, at the solutions do not form a vector space; the form the affine line $u + a$, with a any constant, the integration constant.

Instead, we can consider the equation

$$L(u) = u'' - \frac{f'}{f}u' = 0. \tag{16.1}$$

In addition to u, this also has the constant solutions. In other words, a basis for V_L is given by 1 and u. These are linearly independent because $u' \neq 0$, but $1' = 0$. The symmetry σ of V_L with $\sigma(u) = u + a$, and $\sigma(1) = 1$ of course, acts on this basis of V_L with the matrix

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

The entire symmetry group is isomorphic to the additive group $\mathbb{C}$. To show that σ is a symmetry of the differential field E, check that $\sigma \circ D$ and $D \circ \sigma$ both send a function $g_n u^n + \dots$ to

$$g_n'(u + a)^n + g_n n(u + a)^{n-1} f + \dots.$$

Proposition 16.1. *The function u is transcendental over F.*

Proof. Assume that u satisfies a polynomial equation. Let

$$u^n + g_{n-1}u^{n-1} + \cdots = 0$$

be one such polynomial equation of minimal degree. Differentiating gives

$$nu^{n-1}f + g'_{n-1}u^{n-1} + \cdots = 0$$

By minimality, this polynomial is zero, so that $nf + g'_{n-1} = 0$, from which it follows that f is a derivative. This is a contradiction to our assumption on f. □

Proposition 16.2. *There are no new constants in E: we have $\mathrm{C}_E = \mathrm{C}_F$*

Proof. Assume that $p/q = a$ is constant, where p and q are polynomials in u. Then $p = aq$, and $p' = aq'$, so that p'/q' is also constant, and the polynomial involved are of smaller degree. We can continue this process unless q is a constant, so that $q' = 0$. Then p itself is a constant and we argue as in the preceding proposition: we can write

$$p = g_n u^n + g_{n-1}u^{n-1} + \ldots,$$

and differentiating gives

$$0 = p' = g'_n u^n + g_n n u^{n-1} f + g'_{n-1}u^{n-1} + \ldots.$$

The preceding proposition shows that u is transcendental, so that all coefficients have to vanish. This means that u is constant, and f is a derivative, contradicting our choice of f. □

Example 16.3. For a specific example, take $1/t \in \mathbb{C}(t)$. This is not a derivative in $\mathbb{C}(t)$. (Why?) An integral is $u = \log(t)$, which satisfies the differential equation

$$L(u) = u'' + \frac{1}{t}u' = 0.$$

Example 16.4. There is no element of $\mathbb{C}(t, \exp(-t^2))$ with derivative $\exp(-t^2)$. It follows that the symmetry group of the equation

$$L(u) = u'' + 2tu' = 0$$

is $\mathbb{C}$.

16.2 Exponentials

The exponential function $\exp$ is a solution of the first order linear differential equation

$$L(u) = u' - u = 0.$$

This one has constant coefficients. The solution space is 1–dimensional over $\mathbb{C}$ and consists of all complex multiples of $\exp(t)$, of course. The splitting field over the entire complex plane $U = \mathbb{C}$ is the field which is formed by all fractions $f(t)/g(t)$, where $f(t)$ and $g(t)$ are polynomials in $\exp(t)$ with coefficients that are polynomials in t.

Exercise 57. Is $u = \exp(t)$ algebraic over $\mathbb{C}(t)$? Is the differential field extension it generates, which we described above, isomorphic to $\mathbb{C}(t, u)$ with differential $D(u) = u$? Is there a difference to the field $\mathbb{C}(t)(u)$, or is the situation the same as with $\mathbb{C}[t, u] = \mathbb{C}[t][u]$?

More generally, we consider the linear differential equation

$$L(u) = u' - fu = 0 \tag{16.2}$$

for some function f in F that has an integral (such as $f = 1$ with integral t). Then

$$u = \exp(\int f)$$

is a solution, and if v is any other solution, then v/u is constant. The solution space is the line spanned by u, and the symmetries of the equation are given by the multiplicative group

$$\{u \mapsto au \,|\, a \neq 0 \text{ constant}\} \cong \mathbb{C}^\times.$$

16.3 Algebraic solutions

Again, the situation is best illustrated by an example.

Example 16.5. Let us consider the first order linear differential equation

$$L(u) = u' + \frac{t}{1-t^2}u = 0$$

on the open unit disk. According to the scheme established above, it is solved by

$$u = \exp(\int \frac{-t}{1-t^2}) = \exp\left(\frac{1}{2}\log(1-t^2)\right) = \sqrt{1-t^2}.$$

Note that the integral does not exist in $\mathbb{C}(t)$. Note also that the square root has two 'branches,' and u can denote any one of them. Since $u^2 = 1 - t^2$, the function u is algebraic over $\mathbb{C}(t)$, and the splitting field is the quadratic Galois extension

$$\mathbb{C}(t)[u]/(u^2 + t^2 = 1),$$

with Galois group $\{u \mapsto \pm u\}$ of order 2. The solution space to the differential equation is the line spanned by u, and the symmetry group is contained in the group $\mathrm{GL}_1(\mathbb{C}) = \mathbb{C}^\times$. Which $\sigma(u) = au$ with $a \in \mathbb{C}^\times$ are symmetries? We calculate

$$1 - t^2 = \sigma(1-t^2) = \sigma(w^2) = \sigma(w)^2 = a^2w^2 = a^2(1-t^2)$$

and see that $a = \pm 1$. This shows that the symmetry group of the differential equation agrees with the Galois group of the polynomial equation $X^2 = 1 - t^2$ over the function field $\mathbb{C}(t)$ here.

17 Differential Galois groups

We proceed just as in algebra.

Definition 17.1. If $F \subset E$ is an extension of differential field, so that the differential D on E extends the differential D on F, the *differential Galois group* $\mathrm{Gal}(E|F)$ consists of all automorphisms $E \to E$ of differential fields that are the identity on F.

Remark 17.2. This is a subgroup of the usual group of *all* field automorphisms of $E \to E$ that are the identity on F, which is the usual Galois group studied in algebra. In case we want to make the distinction, we write $\mathrm{Gal}_D(E|F)$.

If $E \subseteq \mathcal{M}(U)$ is the differential splitting field of a linear differential operator L of degree r, then the solutions form an r–dimensional $\mathbb{C}$–vector space V_L. Since E is generated by V_L, the action of the differential Galois group is determined by its effect on V_L, where it acts linearly.

Exercise 58. Verify the details: if σ is in the differential Galois group, then σ maps V_L into itself, and this map is $\mathbb{C}$–linear.

It follows that we have an embedding

$$\mathrm{Gal}_D(E_L|F) \longrightarrow \mathrm{GL}(V_L).$$

Note that we can identify V_L with $\mathbb{C}^r$ by choosing a point u in U and looking at the initial value conditions at u. Consequently, we can identify the group $\mathrm{GL}(V_L)$ with $\mathrm{GL}_r(\mathbb{C})$. This procedure is analoguous to the identification of an algebraic Galois group with a subgroup of the symmetric group via its permutation action on the roots.

Example 17.3. In Section 16.2, we considered the first order differential equation whose solutions are the constant multiples of the exponential function. It follow that the differential Galois groups is a subgroup of $\mathrm{GL}_1(\mathbb{C}) = \mathbb{C}^\times$. It is not hard to see that $\exp(t) \mapsto z\exp(t)$ extends to a field automorphism for each complex number $z \neq 0$, so that the differential Galois group is all of $\mathbb{C}^\times$ in this case.

18 Picard–Vessiot and Liouville extensions

19 The Wronskian

Let u, v be a fundamental system of solution to a second order equation $L(u) = 0$. We write

$$W_L = \begin{pmatrix} u & v \\ u' & v' \end{pmatrix}.$$

The determinant

$$w_L = \det(W_L) = \det \begin{pmatrix} u & v \\ u' & v' \end{pmatrix}$$

is called the *Wronskian* w_L of L. If we change the basis u and v to another basis, the Wrosnkian changes by a non-zero constant. Because u and v are a basis, the Wronskian is not zero. In other words, the Wronskian defines a 1-dimensional line in the function field.

Remark 19.1. The Wronskian is actually nowhere zero.

The problem is that the Wronskian is hard to compute without already knowing a basis u and v of the solution space. If the differential equation is difficult to solve, what then?

We can rewrite the second order equation $L(u) = u'' + cu' + du = 0$ in matrix form

$$A_L \begin{pmatrix} u \\ u' \end{pmatrix} = \begin{pmatrix} u' \\ u'' \end{pmatrix}$$

with the matrix

$$A_L = \begin{pmatrix} 0 & 1 \\ -c & -d \end{pmatrix}$$

that has entries in the differential field.

Proposition 19.2. *The Wronskian satisfies the first order equation*

$$w_L' = \operatorname{tr}(A_L) w_L.$$

Proof. Let $X^\#$ be the adjunct of X, so that $X^\# X = \det(X) = X X^\#$. It is straightforward to verify that $\det(X)' = \operatorname{tr}(X' X^\#)$. We can apply this to $X = W_L$ and get

$$w_L' = \det(W_L)' = \operatorname{tr}(W_L' W_L^\#).$$

Since $W_L' = A_L W_L$' by the differential equation, we get

$$\begin{aligned}
\operatorname{tr}(W_L' W_L^\#) &= \operatorname{tr}((A_L W_L) W_L^\#) \\
&= \operatorname{tr}(A_L (W_L W_L^\#)) \\
&= \operatorname{tr}(A_L \det(W_L)) \\
&= \operatorname{tr}(A_L) \det(W_L) \\
&= \operatorname{tr}(A_L) w_L,
\end{aligned}$$

as claimed. □

Corollary 19.3. *The Wronskian is contained in an extension that is obtained by adjoing an exponential of an integral.*

Proposition 19.4. *Consider the differential field extensions*

$$F \subseteq F(w_L) \subseteq E_L.$$

Then $\operatorname{Gal}(E_L|F(w)) = \operatorname{Gal}(E_L|F) \cap \operatorname{SL}(V_L)$ *inside* $\operatorname{GL}(V_L)$.

Proof. Let σ be in $\operatorname{Gal}(E_L|F(w))$. Then σ acts on the entires of W_L, and

$$\sigma(w_L) = \sigma(\det(W_L)) = \det(\sigma(W_L))$$

because σ is a field automorphism, which commutes with determinants. Let us denote the difference between W_L and $\sigma(W_L)$ by $X = W_L^{-1}\sigma(W_L)$. Then

$$\sigma(W_L) = W_L X$$

shows that X describes the σ action on W_L (and on V_L). This implies

$$\det(X) = \det(W_L)^{-1}\sigma(\det(W_L)),$$

and we see that the element σ fixes the Wronskian w_L if and only if $\det(X) = 1$. □

Corollary 19.5. *If the Wronskian lies in F, then $\mathrm{Gal}(E_L|F) \subset \mathrm{SL}(V_L)$. This applies if w_L is constant.*

Remark 19.6. The Wronskian has an analog in ordinary Galois theory. Given a splitting field E_f of a polynomial f over F, the discriminant Δ_f is in F, and it is a square in E_f. The question is: is it a square in F? In other words: is $\sqrt{\Delta}_f \in F$? We have that $\mathrm{Gal}(E_f|F(\sqrt{\Delta}_f))$ consists if those Galois symmetries that preserve $\sqrt{\Delta}_f$, and this is equal to the subgroup of $\mathrm{Gal}(E_f|F)$ of even permutations of the roots of f.

20 Some special functions

In this section, we shall study a few classes of equations whose solutions can not always be found using the comparatively elementary methods of integration, exponentiation, and algebra: the Airy equation, the Bessel equations, and related equations.

20.1 The Airy equation

The *Airy* operator is the first order linear differential operator given by

$$L(f) = D(f) - tf.$$

This operator is named after the British astronomer Airy (1801–1892). Their solution display turning point phenomena: the change from oscillations to exponential behavior. They also illustrate the Stokes phenomenon: their asymptotic behavior for $|t| \to \infty$ depends on $\arg(t)$ for $t \in \mathbb{C}$. An example of a solution is the Airy function

$$\mathrm{Ai}(t) = \frac{1}{\pi} \int_0^\infty \cos\left(\frac{u^3}{3} + tu\right) \mathrm{d}u.$$

References for the Airy equation in our context: [Kap57, HL11]. Kaplansky has a different sign: $D(f) + tf = 0$. Actually, for most of what is written here, it only matters that the coefficient of f is a polynomial of odd degree.

We can write the Airy equation in matrix form:

$$\begin{pmatrix} 0 & 1 \\ t & 0 \end{pmatrix} \begin{pmatrix} u \\ u' \end{pmatrix} = \begin{pmatrix} u' \\ u'' \end{pmatrix}.$$

We see that the matrix

$$A = \begin{pmatrix} 0 & 1 \\ t & 0 \end{pmatrix}$$

has trace 0, and we'll use that to show that the differential Galois group of the equation lies inside $\mathrm{SL}_2(\mathbb{C})$. Indeed, we have equality, but this is much harder to show. One reason is that we cannot explicitly write down a basis of the solution space. We have to argue indirectly.

Part III: Geometry

21 Affine algebraic groups

The purpose of this section is to explain the geometry of affine algebraic groups such as the general linear groups GL_n.

In differential geometry, manifolds are geometric structures that locally "look and feel" like open subsets $U \subseteq \mathbb{R}^n$. For instance, the general linear group

$$\mathrm{GL}_n(\mathbb{R}) = \{\, A \in \mathrm{M}_n(\mathbb{R}) \mid \det(A) \neq 0 \,\}$$

can be thought of as an open subset of n^2–space.

Similarly, in algebraic geometry, schemes are the fundamental geometric structures that locally "look and feel" like affine schemes. Fortunately for us, the most interesting algebraic groups are affine: the groups GL_n are, and all algebraic subgroups are, too. We can and will make a precise statement soon. This result will justify the approach that we are taking here: it allows us to study the class of linear and affine groups with a comparatively small amount of technology. We'll see that we have already met all concepts that we need!

21.1 The circle

Let us start with an example: the circle. As a manifold, this is not isomorphic to an open subset of n–space, but we will see that it is affine as a scheme because it is given by the solutions of a polynomial equation, namely $x^2 + y^2 = 1$.

As in Part I, we choose a ground ring K that is commutative with unit. The most important case is when K is a field, and it might be a good idea to think of $K = \mathbb{R}$, or better $K = \mathbb{C}$ because that is algebraically closed, at first reading, just as in linear algebra. We let

$$\mathbf{Com}_K$$

denote the category of commutative K–algebras with unit.

The equation of the circle defines a functor

$$C_K \colon \mathbf{Com}_K \longrightarrow \mathbf{Set}$$

that sends a K–algebra A to the set

$$C_K(A) = \{\, (a,b) \in A^2 \mid a^2 + b^2 = 1 \,\}$$

of *A–valued points* of the circle.

It seems natural in this case to choose $K = \mathbb{R}$, but nothing stops us from taking $K = \mathbb{Z}$ here: that's were the polynomial $x^2 + y^2 = 1$ is defined. Of course, one may ask why we have to consider all $\mathbb{R}$–algebras A, and not just $A = \mathbb{R}$. The answer is: $C_K(\mathbb{R})$ is just a set, and knowing the solutions for all other $\mathbb{R}$–algebras A will reveal the geometry of the circle.

We will often leave out K from the notation whenever it is not important. For instance, we have $C_{\mathbb{Z}}(A) = C_{\mathbb{R}}(A)$ for all $\mathbb{R}$–algebras A, but the functor $C_{\mathbb{Z}}$ is defined for all commutative rings with unit, whereas $C_{\mathbb{R}}$ is not defined for $A = \mathbb{Z}$ or $A = \mathbb{Q}$.

21.2 Affine geometry

Of course, there are functors $X \colon \mathbf{Com}_K \to \mathbf{Set}$ that are less interesting than others. The circle has the property that the functor described above is representable:

$$X(A) \cong \mathbf{Com}_{\mathbb{R}}(\mathbb{R}[\mathrm{X},\mathrm{Y}]/(\mathrm{X}^2 + \mathrm{Y}^2 = 1), A).$$

More generally, we have the following definition.

Definition 21.1. An *affine K–scheme* is a representable functor $\mathbf{Com}_K \to \mathbf{Set}$.

Remark 21.2. The more general *K–schemes* are also functors $\mathbf{Com}_K \to \mathbf{Set}$. They are those functors that can be "covered," in a suitable sense, by affine K–schemes. We will not make this precise here, even though it is not very difficult.

It is a formal consequence of the Yoneda lemma that affine K–schemes are the same thing as commutative K–algebras. We will write $\mathcal{O}(X)$ for a commutative K–algebra that represents an affine K–scheme, so that we have

$$X(A) = \mathbf{Com}_{\mathbb{R}}(\mathcal{O}(X), A),$$

and we will write $\mathrm{Spec}(A)$ for the affine K–scheme represented by A, so that we have

$$\mathrm{Spec}(A)(B) = \mathbf{Com}_K(A, B)$$

for all K–algebras B. The only delicate issue worth pointing out is this: a morphism $A \to B$ of commutative K–algebras induces a natural transformation $\mathrm{Spec}(B) \to \mathrm{Spec}(A)$ of functors: the direction of the arrows is reversed. We have already seen that before, in Section 9.5. For that reason, we have to say:

Proposition 21.3. *The category of affine K–schemes is equivalent to the opposite of the category of commutative K–algebras.*

Example 21.4. The *affine n–space* (over K) is the functor

$$\mathbb{A}^n_K \colon \mathbf{Com}_K \to \mathbf{Set},\ A \mapsto A^n.$$

It is representable by the polynomial ring $K[\mathrm{X}_1, \ldots, \mathrm{X}_n]$ in n variables. We can write this as $\mathbb{A}^n_K = \mathrm{Spec}(K[\mathrm{X}_1, \ldots, \mathrm{X}_n])$ and $K[\mathrm{X}_1, \ldots, \mathrm{X}_n] = \mathcal{O}(\mathbb{A}^n_K)$.

Exercise 59. Let X be an affine K–scheme. Show that there is a natural bijection between $\mathcal{O}(X)$ and the set of all morphisms $X \to \mathbb{A}^1_K$ of affine K–schemes. In this way, the elements of the commutative K–algebra $\mathcal{O}(X)$ can be thought of as functions on X with values in the affine line.

A surjective morphism $A \to B$ of commutative K–algebras gives rise to a morphism $\mathrm{Spec}(B) \to \mathrm{Spec}(A)$ that is injective when evaluated on each commutative K–algebra C: a morphism $B \to C$ is essentially the same thing as a morphism $A \to C$ that sends the kernel I of $A \to B$ to zero. We will refer to $\mathrm{Spec}(B) \to \mathrm{Spec}(A)$ as the *closed embedding* defined by the ideal $I \subseteq A$.

Example 21.5. The special linear group SL_n comes with a closed embedding into $\mathbb{A}^{n^2}$:

$$\mathrm{SL}_n(A) = \{\, \mathrm{X}_{i,j} \mid \det(\mathrm{X}_{i,j}) = 1 \,\}.$$

This functor is an affine scheme:

$$\mathcal{O}(\mathrm{SL}_n) = K[\mathrm{X}_{i,j}]/(\det(\mathrm{X}_{i,j}) = 1).$$

The general linear group GL_n is also an affine K–scheme: introduce another variable Y which plays the role of an inverse of $\det(\mathrm{X}_{i,j})$. Then we can write

$$\mathrm{GL}_n(A) = \{\, \mathrm{X}_{i,j} \mid \det(\mathrm{X}_{i,j}) \text{ invertible} \,\}.$$

and

$$\mathcal{O}(\mathrm{GL}_n) = K[\mathrm{X}_{i,j}, Y]/(\det(\mathrm{X}_{i,j})Y = 1).$$

The closed embedding $\mathrm{SL}_n \to \mathrm{GL}_n$ corresponds to the ideal $Y = 1$ in $\mathcal{O}(\mathrm{GL}_n)$.

21.3 Affine group schemes

The definition of an affine group scheme (over K) is now easy:

Definition 21.6. An *affine group scheme* (over K) is a functor

$$G\colon \mathbf{Com}_K \longrightarrow \mathbf{Grp}$$

such that the composition $\mathbf{Com}_K \to \mathbf{Grp} \to \mathbf{Set}$ with the forgetful functor is an affine K–scheme.

This is a precise version of Definition 1.5.

Example 21.7. It is clear that GL_n is an affine group scheme: the value on an commutative K–algebra is the group $\mathrm{GL}_n(B)$ of invertible matrices with entries in B. This is the group scheme of K–automorphisms of the vector space K^n. What if we want to work in a co-ordinate invariant way with a K–vector space V? We

want an affine group scheme $\mathrm{GL}(V)$, and we have to say what its value on a commutative K–algebra B is. For this we note that $\mathrm{GL}_n(B)$ is the group of B–automorphisms of the free B–module

$$B^n = (B \otimes_K K)^n = B \otimes_K (K^n).$$

This suggests to set

$$\mathrm{GL}(V)(B) = \mathrm{GL}(B \otimes_K V),$$

where GL on the right hand side refers to the B–liner automorphisms of the free B–module $B \otimes_K V$.

Example 21.8. It is straightforward to generalizing the preceding example to 'structured vector spaces:' if $(V, \star)$ is any K–algebra (commutative, merely associative, Lie, or other), then $\mathrm{Aut}(V, \star)$ will be the affine (!) subgroup scheme of $\mathrm{GL}(V)$ that takes B to the subgroup $\mathrm{Aut}(B \otimes_K V, B \otimes_K \star)$ of $\mathrm{GL}(B \otimes_K V)$.

Proposition 21.9. *The category of affine group schemes over K is equivalent to the opposite of the category of commutative Hopf algebras over K.*

Proof. The correspondence is the same as in Proposition 21.3: Given G, the K–algebra $H = \mathcal{O}(G)$ is a Hopf algebra. For instance, the comultiplication is given by morphism

$$H = \mathcal{O}(G) \longrightarrow \mathcal{O}(G \times G) \cong \mathcal{O}(G) \otimes \mathcal{O}(G) = H \otimes H$$

induced by the multiplication $G \times G \to G$ of G. In that calculation, we used the isomorphism $\mathcal{O}(G \times G) \cong \mathcal{O}(G) \otimes \mathcal{O}(G)$ comes from the fact that $\mathcal{O}$, as a contravariant functor, turns product into sums, and $\otimes$ is the sum in the category of commutative K–algebras.

Conversely, if H is a Hopf K–algebra, then $G = \mathrm{Spec}(H)$ is an affine group scheme over K: the group structure on

$$G(A) = \mathbf{Com}_K(H, A)$$

is given as follows: if $g, h\colon H \to A$ are two morphisms of K–algebras, then the composite

$$H \xrightarrow{\Delta_H} H \otimes H \xrightarrow{g\otimes h} A \otimes A \xrightarrow{m_A} A$$

is their product. □

Exercise 60. Fill in the remaining details in the preceding proof.

Example 21.10. The additive group $\mathbb{G}_{\mathrm{a}}$ is the functor

$$\mathbb{G}_{\mathrm{a}}(A) = A,$$

where the values are groups under addition. This functor is represented by the polynomial K–algebra $K[\mathrm{T}]$, and the structure maps are as in Section 9.

Example 21.11. The multiplicative group $\mathbb{G}_{\mathrm{m}}$ is the functor

$$\mathbb{G}_{\mathrm{m}}(A) = A^{\times},$$

where the values are groups of invertible elements under multiplication. This functor is represented by the K–algebra $K[\mathrm{T}^{\pm}]$ of Laurent polynomials, and the structure maps are as in Section 9. The n–th roots of unity define a subgroup K–scheme $\mu_n$representable by $\mathcal{O}(\mu_n) = K[\mathrm{T}]/(T^n - 1)$.

Exercise 61. Check that the formula

$$(a, b) \cdot (c, d) = (ac - bd, bc + ad)$$

can be used to turn the real circle $C_{\mathbb{R}}$ into an affine group scheme over $\mathbb{R}$. What are the co-multiplication, co-unit, and antipode on $\mathbb{R}[\mathrm{X}, \mathrm{Y}]/(\mathrm{X}^2 + \mathrm{Y}^2 = 1)$?

Exercise 62. Let K be a field of positive characteristic p. Show that

$$\alpha_p\colon A \mapsto \{a \in A \,|\, a^p = a\}$$

is an affine group scheme over K. What is the ring $\mathcal{O}(\alpha_p)$ of functions on it? What are the co-multiplication, co-unit, and antipode on it?

22 Linear algebraic groups

Example 22.1. It is very instructive to see what we get for the general linear group schemes $G = \mathrm{GL}_n$. Recall that

$$H = \mathcal{O}(\mathrm{GL}_n) = K[\mathrm{X}_{i,j}, \mathrm{Y}]/(\det(\mathrm{X}_{i,j})\mathrm{Y} = 1).$$

The co-multiplication is given by

$$\Delta(\mathrm{X}_{i,k}) = \sum_j \mathrm{X}_{i,j} \otimes \mathrm{X}_{j,k}.$$

Of course, this is a reflection of the usual formula for matrix multiplication. The co-unit is given by

$$\varepsilon(\mathrm{X}_{i,k}) = \delta_{i,k}$$

because the entries of the identity matrix are given by Kronecker's δ. Finally, the formula ("Cramer's rule") for the antipode $S(\mathrm{X}_{i,k}) = Y_{i,k}$ is determined by the entries $Y_{i,k} \in H$ of the inverse matrix of $(\mathrm{X}_{i,k})$ in $\mathrm{GL}_n(H)$. Note that $\mathrm{Y} = \det(Y_{i,k})$ for this inverse. Note that $\mathrm{GL}_1 = \mathbb{G}_\mathrm{m}$.

Exercise 63. The value of Δ, ε, and S on Y is determined by these formulas. Can we make it explicit?

A *linear* algebraic group is a closed subgroup scheme of some GL_n.

Since GL_n is an affine scheme of finite type, so is every closed subscheme. Therefore, all linear algebraic groups are affine group schemes of finite type. The converse is also true:

Theorem 22.2. *Every affine group scheme of finite type over a field is linear.*

Proof. Let us take an affine group scheme G of finite type over a field K and show that it can be embedded as a closed subscheme of some GL_n. Set $H = \mathcal{O}(G)$ be the associated Hopf algebra. By assumption H is a finitely generated K–algebra.

We claim that every $h \in H$ is contained in a finite-dimensional subspace $V_h \subseteq H$ that is a 'co-ideal' in the sense that $\Delta(V_h) \subseteq V_h \otimes H$. To prove the claim, choose a basis b_i of H and write

$$\Delta(h) = \sum_i v_i \otimes b_i,$$

with finitely many v_i, and

$$\Delta(b_i) = \sum_{j,k} \lambda^i_{j,k} b_j \otimes b_k.$$

By co-associativity, we have

$$\sum_i \Delta(v_i) \otimes b_i = (\Delta \otimes \mathrm{id})\Delta(h) = (\mathrm{id} \otimes \Delta)\Delta(h) = \sum_{i,j,k} v_i \otimes \lambda^i_{j,k} b_j \otimes b_k$$

Comparing coefficients in front of b_k, we get

$$\Delta(v_k) = \sum_{i,j} v_i \otimes \lambda^i_{j,k} b_j.$$

This shows that the span of h and the finitely many v_i is a subspace V_h with the desired properties. This proves the claim.

If $h_1, \ldots, h_m$ are K–algebra generators of H, then $W = \sum_i V(h_i)$ is a finite-dimensional co-ideal of H that contains them. We choose a K–basis $w_1, \ldots, w_n$ of W and write

$$\Delta(w_j) = \sum_i w_i \otimes a_{i,j} \tag{22.1}$$

for uniquely determined $a_{i,j} \in H$. Then we define

$$K[\mathrm{X}_{i,j}, \mathrm{Y}]/(\det(\mathrm{X}_{i,j})\mathrm{Y} - 1) \longrightarrow H,\ \mathrm{X}_{i,j} \longmapsto a_{i,j}.$$

We leave some checking to the exercises, but we will check that it is surjective. Since it is a K–algebra morphism by construction, it suffices to show that it hits the generators h_i. These lie inside W, which is spanned by the w_j. So it suffices to show that the w_j are hit. The co-unit axiom gives

$$w_j = (\varepsilon \otimes \mathrm{id})\Delta(w_j) = (\varepsilon \otimes \mathrm{id}) \sum_i w_i \otimes a_{i,j} = \sum_i \varepsilon(w_i) a_{i,j},$$

and the $a_{i,j}$ are in the image by construction. □

In the course of the preceding proof, we have seen more generally:

Proposition 22.3. *Every co-module over a co-algebra over a field is the union of its finite-dimensional sub-co-modules.*

Exercise 64. Show that the $a_{i,j}$ defined in (22.1) satisfy the same formulas for the structure maps Δ, ε, and S as the algebra generators for $\mathcal{O}(\mathrm{GL}_n)$. These are, of course, just the formulas for matrix multiplication from linear algebra in another guise. For instance, to find the formula for the co-multiplication, use co-associtivity and the definition (22.1). On the one hand, we have

$$\begin{aligned}(\Delta \otimes \mathrm{id})\Delta(w_k) &= (\Delta \otimes \mathrm{id}) \sum_j w_j \otimes a_{j,k} \\ &= \sum_i \Delta(w_j) \otimes a_{j,k} \\ &= \sum_{i,j} w_i \otimes a_{i,j} \otimes a_{j,k}\end{aligned}$$

On the other hand, we have

$$\begin{aligned}(\mathrm{id} \otimes \Delta)\Delta(w_k) &= (\mathrm{id} \otimes \Delta) \sum_i w_i \otimes a_{i,k} \\ &= \sum_i w_i \otimes \Delta(a_{i,k})\end{aligned}$$

Comparing coefficients, this gives

$$\Delta(a_{i,k}) = \sum_j a_{i,j} \otimes a_{j,k},$$

as desired.

The notion of an affine algebraic group stresses the independence from any embedding. The notion of a linear algebraic group stresses the concreteness of the object.

Exercise 65. Go back to Section 9.5 and check that we have already seen there that every finite group G defines a linear/affine algebraic group through the ring $\mathcal{O}(G)$ of functions $G \to K$.

23 Tangent spaces and Lie algebras

Tangent spaces are supposed to linearize a geometric object in a given point. If the geometric object is a scheme, we can assume that it is affine because locally around any given point it is affine. But what is a 'point' of an affine K–scheme X? If B is a K–algebra, we have called the elements of

$$X(B) \cong \mathbf{Com}_K(\mathcal{O}(X), B)$$

the B–valued points of X. A preferred choice for the K–algebra B is of course the initial one, namely $B = K$. A K–valued point of X is a morphism $\xi\colon \mathcal{O}(X) \to K$ of K–algebras. Note that there is also a unique morphism $K \to \mathcal{O}(X)$ of K–algebras, the unit.

There is an affine K–scheme T that is, in geometric terms, the "universal point that carries a tangent vector." It is given as $T = \mathrm{Spec}(K[t]/t^2)$. It is obvious that T has a unique K–valued point, so that $T(K) = \star$. A morphism $T \to X$ of affine K–schemes, therefore, determines a K–point ξ of X, the composition

$$\mathcal{O}(X) \longrightarrow T \longrightarrow K.$$

Definition 23.1. A tangent vector of an affine K–scheme X at a K–point ξ is a K–morphism $T \to X$ that maps to the K–point ξ. We write $T_\xi X$ for the set of tangent vectors of X at ξ.

In algebraic terms, a tangent vector to an affine K–scheme X at a K–point ξ is given by a morphism

$$\mathcal{O}(X) \longrightarrow \mathcal{O}(T) = K[t]/t^2$$

which lifts ξ: it gives $\xi\colon \mathcal{O}(X) \to K$ under $t \mapsto 0$. Another way of saying this: the tangent space $T_\xi X$, as a set, is the pre-image of ξ under the map

$$X(K[t]/t^2) \longrightarrow X(K) \tag{23.1}$$

induced by the map $t \mapsto 0$. Perhaps this also suggests to think of $X(K[t]/t^2)$ as the set of K–points of the tangent bundle of X.

At this point it would be good to revisit Exercise 55.

Proposition 23.2. *There is a canonical bijection between $T_\xi X$ and the set of K–derivations*

$$\delta\colon \mathcal{O}(X) \longrightarrow K$$

of $\mathcal{O}(X)$, where K is an $\mathcal{O}(X)$–module via ξ:

$$\delta(fg) = \delta(f)\xi(g) + \xi(f)\delta(g).$$

Remark 23.3. It might be helpful to read $\xi(f)$ as $f(\xi)$ because, after all, this is the evaluation of the function f at the point ξ. Thankfully, we are still doing math and not Python, where we would have to write $\xi.f()$, or?

Proof. Given such a derivation δ, define

$$\mathcal{O}(X) \longrightarrow K[t]/t^2$$

by $f \mapsto \xi(f) + t\delta(f)$. This is a K–morphism as described above. Conversely, the coefficient in front of (or behind) t is always a K–derivation as above. □

Corollary 23.4. *The set $T_\xi X$ of tangent vectors at ξ has a canonical K–vector space structure.*

Example 23.5. Let us look at the circle $C_\mathbb{R}$ again. Given an $\mathbb{R}$–point $(x, y) \in \mathbb{R}^2$ with $x^2 + y^2 = 1$, a tangent vector to it is a K–morphism

$$\mathbb{R}[\mathrm{X}, \mathrm{Y}]/(\mathrm{X}^2 + \mathrm{Y}^2 - 1) \longrightarrow \mathbb{R}[t]/t^2$$

that reduces to (x, y) modulo t. Therefore, it is given as

$$\mathrm{X} \longmapsto x + tv$$
$$\mathrm{Y} \longmapsto y + tw$$

for suitable (v, w). Which? We need

$$1 = (x + tv)^2 + (y + tw)^2 = 1 + 2t(xv + yw),$$

and this is equivalent to (v, w) being orthogonal to (x, y), as expected:

$$T_{(x,y)}C_\mathbb{R} = \{\, (v, w) \in \mathbb{R}^2 \mid (v, w) \perp (x, y) \,\}.$$

Example 23.6. The tangent space of GL_n at the identify matrix is the space $\mathrm{gl}_n(K)$ of all (n,n)–matrices with entries in K. As the notation suggests, we'll see next that this is indeed a Lie algebra over K. The tangent space of SL_n at the identify matrix is the space sl_n of all (n,n)–matrices with entries in K that have trace 0. Let us verify this for SL_2: the matrix

$$\begin{pmatrix} 1+ta & tb \\ tc & 1+td \end{pmatrix}$$

has determinant

$$(1+ta)(1+td) - (tb)(tc) = 1 + t(a+d) \in K[t]/(t^2).$$

Therefore, it lies inside $\mathrm{SL}_2(K[t]/(t^2))$ if and only if $a+d=0$. For $\mathrm{GL}_2(K[t]/(t^2))$, we don't need any condition because $1+t(a+d)$ is always invertible in $K[t]/(t^2)$ with inverse $1-t(a+d)$. The general case follows from Leibniz' formula for the determinant.

Exercise 66. What's the tangent space of the 'nodal' singularity

$$\{\, (x,y) \in \mathbb{A}^2 \,|\, y^2 = x^3 + x^2 \,\}$$

at the origin $(0,0) \in \mathbb{A}^2$? What's the tangent space of the 'cusp' singularity

$$\{\, (x,y) \in \mathbb{A}^2 \,|\, y^2 = x^3 \,\}$$

at the origin $(0,0) \in \mathbb{A}^2$? What's the tangent space of the elliptic curve

$$\{\, (x,y) \in \mathbb{A}^2 \,|\, y^2 = x^3 - x \,\}$$

at the origin $(0,0) \in \mathbb{A}^2$?

23.1 Lie algebras of affine group schemes

It is always possible to associate a Lie algebra to *any* affine scheme X over K in a geometrically meaningful way. Indeed, if $A = \mathcal{O}(X)$ is the ring of functions on A, we can think of the K–vector space

$$\mathrm{Der}_K(\mathcal{O}(X), \mathcal{O}(X))$$

as the Lie algebra of vector fields on X, as in Example 6.8. This is a Lie algebra as we have seen more generally in Proposition 6.7.

If $X = G$ is an affine *group* scheme, then something special happens. We can use the group multiplication to move any point to any other point in a systematic way, and this 'trivializes the tangent bundle:' every tangent space T_gG is identified with the tangent space T_eG at the neutral element e, and a vector field is the same as a map $G \to T_eG$. Turning this argument around, we expect that T_eG reappears as the K–vector space of 'G–invariant' vector fields on G. We make this precise now and use it to endow T_eG with a Lie algebra structure.

Remark 23.7. We remark that the neutral element in $G(K) = \mathbf{Com}_K(\mathcal{O}(G), K)$ is ε, the co-unit of the Hopf algebra $\mathcal{O}(G)$ over K. For that reason, it might have been better to write $T_\varepsilon G$ above. On the other hand, that paragraph was meant as a motivation only, and precision was not what it was aiming at.

23.2 Invariant operators

Let G be an affine group scheme over a field K, corresponding to the commutative Hopf algebra $H = \mathcal{O}(G)$ over K.

Definition 23.8. We say that an operator $D \in \mathrm{End}_K(\mathcal{O}(G))$ is *G–invariant* if the diagram

$$\begin{array}{ccc} H & \xrightarrow{\quad D \quad} & H \\ {\scriptstyle\Delta}\downarrow & & \downarrow{\scriptstyle\Delta} \\ H \otimes H & \xrightarrow[\mathrm{id}\otimes D]{} & H \otimes H \end{array}$$

commutes.

Remark 23.9. Strictly speaking, this is left or right invariant, and the other direction is defined using $D \otimes \mathrm{id}$ in the diagram above. We'll use only this version, so we don't have to bother.

Exercise 67. Show that the subset $\mathrm{Inv}_K(\mathcal{O}(G)) \leqslant \mathrm{End}_K(\mathcal{O}(G))$ of G–invariant operators is a K–subalgebra.

23.3 Invariant derivations

Proposition 23.10. *Let G be an affine group scheme over a field K. There are canonical isomorphisms between the following three K–vector spaces*
(1) $\mathrm{Inv}_K(\mathcal{O}(G)) \cap \mathrm{Der}_K(\mathcal{O}(G), \mathcal{O}(G))$,
(2) $\mathrm{Der}_K(\mathcal{O}(G), K)$, *and*
(3) $\mathrm{Ker}(G(K[t]/t^2) \to G(K)) = T_\varepsilon G$.

Proof. We have already seen a canonical bijection between (2) and (3) earlier in this section: the tangent space (2) at ε is the pre-image of ε under the map (23.1), and in our situation this pre-image is the usual kernel (3).

If $D\colon \mathcal{O}(G) \to \mathcal{O}(G)$ is a derivation, the composition

$$\delta = \varepsilon D\colon \mathcal{O}(G) \xrightarrow{D} \mathcal{O}(G) \xrightarrow{\varepsilon} K$$

is a derivation. (Check this!) This gives a map from (1) to (2).

Using the co-unit, we can write $D = (\mathrm{id} \otimes \varepsilon)\Delta D$. Since D is G–invariant, this equals

$$D = (\mathrm{id} \otimes \varepsilon)\Delta D = (\mathrm{id} \otimes \varepsilon)(\mathrm{id} \otimes D)\Delta = (\mathrm{id} \otimes \varepsilon D)\Delta = (\mathrm{id} \otimes \delta)\Delta.$$

This shows that D is determined by its image δ, and the map is injective. Moreover, we can use this formula to define the inverse map from (2) to (1): given δ, we set

$$D = (\mathrm{id} \otimes \delta)\Delta\colon \mathcal{O}(G) \xrightarrow{\Delta} \mathcal{O}(G) \otimes \mathcal{O}(G) \xrightarrow{\mathrm{id} \otimes \delta} \mathcal{O}(G) \otimes K \cong \mathcal{O}(G).$$

This D is a derivation. (Check this!) Let us check that it is G–invariant. On the one hand, we have

$$\Delta D = \Delta(\mathrm{id} \otimes \delta)\Delta.$$

On the other hand, we can use co-associativity to get

$$\begin{aligned}(\mathrm{id} \otimes D)\Delta &= (\mathrm{id} \otimes \mathrm{id} \otimes \delta)(\mathrm{id} \otimes \Delta)\Delta \\ &= (\mathrm{id} \otimes \mathrm{id} \otimes \delta)(\Delta \otimes \mathrm{id})\Delta \\ &= \Delta(\mathrm{id} \otimes \delta)\Delta,\end{aligned}$$

which is the same, showing G–invariance.

The maps are inverse to each other: we have $\varepsilon(\mathrm{id}\otimes\delta)\Delta = \delta$ and $(\mathrm{id}\otimes\varepsilon D)\Delta = D$. □

Exercise 68. Check the unproven claims in the preceding proof.

23.4 The Lie algebra structure

If G is an affine group scheme, we define its Lie algebra $\mathrm{Lie}(G)$ to be any of the three vector spaces in Proposition 23.10. This is obviously a Lie algebra in version (1). Indeed, we already know that the commutator $[D_1, D_2] = D_1D_2 - D_2D_1$ of two G–invariant derivations $D_j\colon \mathcal{O}(G) \to \mathcal{O}(G)$ is again a G–invariant derivation (Exercise 67).

How can we describe the Lie bracket in the version (2) of the Lie algebra? Given two derivations $\delta_j\colon \mathcal{O}(G) \to K$, we can use the isomorphisms in Proposition 23.10 to transport the structure to version (2). First, form $D_j = (\mathrm{id}\otimes\delta_j)\Delta$, then form the commutator $[D_1, D_2]$ as above, and finally set

$$[\delta_1, \delta_2] = \varepsilon[D_1, D_2] = \varepsilon[(\mathrm{id}\otimes\delta_1)\Delta, (\mathrm{id}\otimes\delta_2)\Delta].$$

An easy check reveals that this equals

$$[\delta_1, \delta_2] = (\delta_1 \otimes \delta_2 - \delta_2 \otimes \delta_1)\Delta :$$

both expressions give

$$[\delta_1, \delta_2](f) = \sum_j \Big(\delta_1(g_j)\delta_2(h_j) - \delta_2(g_j)\delta_1(h_j)\Big)$$

on a function f such that $\Delta(f) = \sum_j g_j \otimes h_j$.

How can we describe the Lie bracket in the version (3) of the Lie algebra? Two derivations $\delta_j\colon \mathcal{O}(G) \to K$ correspond to two elements $g_j = \varepsilon + t\delta_j$ in the group

$$G(K[t]/(t^2)) = \mathbf{Com}_K(\mathcal{O}(G), K[t]/(t^2)),$$

with their inverses given by $g_j^{-1} = \varepsilon - t\delta_j$. Recall that the co-unit $\varepsilon\colon \mathcal{O}(G) \to K$ of the Hopf algebra $\mathcal{O}(G)$ is the unit in the group $G(K)$. We'll move these two elements into the group $G(K[u,v]/(u^2,v^2))$ using $t \mapsto u$ and $t \mapsto v$, respectively. We compute their commutator in the sense of group theory:

$$\begin{aligned}&(\varepsilon + u\delta_1)(\varepsilon + v\delta_2)(\varepsilon + u\delta_1)^{-1}(\varepsilon + v\delta_2)^{-1}\\ =&(\varepsilon + u\delta_1 + v\delta_2 + uv\delta_1\delta_2)(\varepsilon - u\delta_1 - v\delta_2 + uv\delta_1\delta_2)\\ =&\varepsilon + uv[\delta_1, \delta_2].\end{aligned}$$

This is the image of $\varepsilon + t[\delta_1, \delta_2]$ under the morphism

$$K[t]/(t^2) \longrightarrow K[u,v]/(u^2,v^2),\ t \mapsto u + v$$

because $(u+v)^2 = uv$ in $K[u,v]/(u^2,v^2)$.

Example 23.11. If G is abelian, then the Lie bracket on $\mathrm{Lie}(G)$ is zero.

Example 23.12. Let $b\colon V \otimes V \to V$ a bilinear map. We can think of this as a multiplication of an algebra of sorts. We are interested in the automorphism group $\mathrm{Aut}(V,b)$ of this structure. Recall that automorphism only depend on the data of an algebraic structure, not on the axioms. Therefore, the group $\mathrm{Aut}(V,b)$ is the closed subgroup of $\mathrm{GL}(V)$ consisting of the $g\colon V \to V$ such that

$$gb(v,w) = b(gv, gw),$$

regardless of the axioms that b satisfies or not. The Lie algebra to $\mathrm{Aut}(V,b)$ consists of those $D\colon V \to V$ such that the element $g = \mathrm{id} + tD$ in $\mathrm{GL}(V \otimes K[t]/t^2)$ preserves b:

$$(\mathrm{id} + tD)b(v,w) = b(v,w) + tDb(v,w)$$

and

$$b((\mathrm{id} + tD)v, (\mathrm{id} + tD)w) = b(v,w) + t(b(Dv,w) + b(v,Dw))$$

show that this is equivalent to D being a derivation:

$$Db(v,w) = b(Dv,w) + b(v,Dw).$$

In summary, we have

$$\mathrm{Lie}\,\mathrm{Aut}(V,b) = \mathrm{Der}(V,b). \tag{23.2}$$

Exercise 69. Verify that a closed embedding $G \to H$ of affine group schemes induces an injective morphism $\mathrm{Lie}(G) \to \mathrm{Lie}(H)$ of Lie algebras.

24 Representations of affine algebraic groups

Let new G be an affine group scheme over the field K with commutative Hopf algebra $H = \mathcal{O}(G)$ over K. A finite-dimensional representation of G is a finite-dimensional K–vector space V together with a co-action Δ_V that is co-associative and co-unital. It can be either a right co-action $V \to V \to H$ or a left co-action $V \to H \otimes V$, but we should stick to one convention.

24.1 Tensor products and trivial representations

A special feature of the category of representations of a Hopf algebra (in contrast to a mere co-algebra) is that we have tensor products

$$\begin{array}{ccc} V \otimes W & \xrightarrow{\Delta_{V\otimes W}} & V \otimes W \otimes H \\ {\scriptstyle \Delta_V \otimes \Delta_W}\downarrow & & \uparrow{\scriptstyle \mathrm{id}_V \otimes \mathrm{id}_W \otimes m} \\ V \otimes H \otimes W \otimes H & \xrightarrow{\cong} & V \otimes W \otimes H \otimes H \end{array}$$

of representations and trivial representations

$$K \xrightarrow{e} H \cong K \otimes H.$$

Actually there are also dual representations (using the antipode), but we will not use them.

24.2 Representation of the multiplicative group

Let $G = \mathbb{G}_{\mathrm{m}}$ be the multiplicative group with $H = K[\mathrm{T}^{\pm 1}]$. A representation is given by a co-action

$$V \longrightarrow V \otimes K[\mathrm{T}^{\pm 1}].$$

For instance, if we choose an $n \in \mathbb{Z}$, then $v \mapsto v \otimes \mathrm{T}^n$ is a representation. It turns out the all $\mathbb{G}_\mathrm{m}$–representations are direct sums of those. Indeed if V is a $\mathbb{G}_\mathrm{m}$–representation, then

$$V_m = \{\, v \in V \mid \Delta_V(v) = v \otimes \mathrm{T}^n \,\}$$

is a sub-representation, and $V_m \cap V_n = 0$ for $m \neq n$. Given v in V, we can write

$$\Delta(v) = \sum_n v_n \otimes \mathrm{T}^n$$

with finitely many v_n non-zero. Applying the co-unit shows $v = \sum_n v_n$. It follows from co-associativity of Δ_V that $v_n \in V_n$. This shows that the V_n span V:

$$V = \bigoplus_n V_n.$$

In summary, we see that $\mathbb{G}_\mathrm{m}$–representations are essentially the same as graded K–vector spaces.

More generally, representations of tori

$$\mathbb{G}_\mathrm{m}^r = \underbrace{\mathbb{G}_\mathrm{m} \times \cdots \times \mathbb{G}_\mathrm{m}}_{r}$$

are essentially the same as multi-graded vector spaces.

24.3 Representation of the additive group

Let $G = \mathbb{G}_\mathrm{a}$ be the additive group over a field K or characteristic 0. We have already seen the $\mathbb{G}_\mathrm{a}$–representation

$$\mathbb{G}_\mathrm{a} \longrightarrow \mathrm{GL}_2,\ t \mapsto \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Note that the right hand side is the exponential of the nilpotent matrix

$$\begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix}.$$

More generally, a finite–dimensional $\mathbb{G}_{\mathrm{a}}$–representation V is essentially the same as a nilpotent endomorphism $f\colon V \to V$. Indeed, a representation is given by

$$\Delta_V\colon V \longrightarrow V \otimes K[\mathrm{T}],\ v \mapsto \sum_{n\geqslant 0} \delta_0(v) \otimes \mathrm{T}^n.$$

This determines a sequence $(\,\delta_n \,|\, n \geqslant 0\,)$ of endomorphisms, eventually zero. The co-unit axiom implies δ_0, and co-associativity and $\Delta(\mathrm{T}^n) = \Delta(\mathrm{T})^n = (\mathrm{T}\otimes 1 + 1\otimes \mathrm{T})^n$ imply

$$\delta_m\delta_n = \binom{m+n}{n}\delta_{m+n}.$$

Setting $f = \delta_1$, this implies

$$\Delta_V(v) = \sum_n f^n(v) \otimes \frac{\mathrm{T}^n}{n!}.$$

24.4 Some general features

Let G be an affine algebraic group. Recall the general fact, stated as Proposition 22.3, that every co-module V is the union of its finite-dimensional sub-representations. This applies, in particular, to the co-module H and its finite sums.

Proposition 24.1. *Every finite-dimensional G–representation is a sub-representation of a finite sum of H's.*

Proof. If V is an n–dimensional G-representation, then $V \otimes H$ is a co-module via

$$V \otimes H \xrightarrow{V\otimes\Delta} V \otimes H \otimes H,$$

isomorphic to H^n. (This is not the co-diagomal action unless V is trivial to begin with.) The co-action

$$\Delta_V\colon V \longrightarrow V \otimes H$$

is a morphism of G–co-modules by co-associativity of Δ_V. It is injective by the co-unitality of Δ_V. □

25 Tannaka duality

In this section, we see how to recover an affine group scheme from its category of finite-dimensional representations. We start be

25.1 Actions of finite groups

Let G be a finite group. Let $\mathbf{Set}_{\mathrm{fin}}^G$ denote the category of finite G–sets. The objects (X, ρ_X) are finite sets X together with an action of the group G in the form of a morphism $\rho_X \colon G \to \Sigma_X$ of groups into the symmetric group Σ_X of all permutations of X. A morphisms of G–sets is a map $f \colon X \to Y$ that is compatible with the G–actions: $f(gx) = gf(x)$. Let

$$\omega \colon \mathbf{Set}_{\mathrm{fin}}^G \longrightarrow \mathbf{Set}_{\mathrm{fin}}$$

be the forgetful functor $(X, \rho_X) \mapsto X$. Let $\mathrm{Aut}(\omega)$ be the group of all natural isomorphisms $\omega \to \omega$. An element σ in the group $\mathrm{Aut}(\omega)$ is a family

$$\sigma(X, \rho_X) \colon \omega(X, \rho_X) \longrightarrow \omega(X, \rho_X) = X$$

of bijections such that for each morphisms $f \colon (X, \rho_X) \to (Y, \rho_Y)$ the diagram

$$\begin{array}{ccc} \omega(X, \rho_X) & \xrightarrow{\sigma(X,\rho_X)} & \omega(X, \rho_X) \\ {\scriptstyle \omega(f)}\downarrow & & \downarrow{\scriptstyle \omega(f)} \\ \omega(Y, \rho_Y) & \xrightarrow[\sigma(Y,\rho_Y)]{} & \omega(Y, \rho_Y) \end{array}$$

commutes. Let $\mathrm{Aut}_{\amalg}(\omega)$ be the subgroup of all the σ that are compatible with disjoint unions. This just means

$$\sigma(X \amalg Y, \rho_{X \amalg Y}) = \sigma(X, \rho_X) \amalg \sigma(Y, \rho_Y).$$

Since there is a unique action τ of G on the empty set $\emptyset$, the canonical requirement $\sigma(\emptyset, \tau) = \mathrm{id}_\emptyset$ is automatic.

Theorem 25.1. *There is a canonical isomorphism*

$$G \longrightarrow \mathrm{Aut}_{\amalg}(\omega)$$

of groups.

Proof. First, we need to product a morphism. Given an element g in G, we set

$$\varphi_g(X, \rho_X) = \rho_X(g) \colon X \to X.$$

In other words, we use that the set underlying a G–set naturally comes with bijections given the element of G. After all, that is what a G–set is. Naturality follows from the fact that the f's are G–maps. This given the morphism φ of groups.

This morphism φ is injective: if $\varphi_g(X, \rho_X) = \mathrm{id}_X$ for all finite G–sets, then we have an element g that acts trivially on all finite G–sets. Since only $g = e$ acts trivially on the finite G–set $X = G$, this shows injectivity.

As for surjectivity, let $\sigma \in \mathrm{Aut}_{\amalg}(\omega)$ be any element. Then $\sigma(G) \colon G \to G$ sends e to some element g in G. We claim that $\sigma = \varphi_g$. If G/H is an orbit, and $xH \in G/H$ any element, there is a unique morphism $f \colon G \to G/H$ of G-sets that maps e to xH. Naturality with respect to this G–map shows that both $\varphi_g(G/H)$ and $\sigma(G/H)$ send xH to gxH. A general finite G–set is the finite disjoint union of its orbits, and the equality $\sigma = \varphi_g$ follows from the compatibility of both with respect to disjoint unions. □

Remark 25.2. The natural generalization of the preceding result is not from finite groups to discrete groups but from finite groups to 'pro-finite' groups.

25.2 Tannaka duality for affine group schemes

Tannaka proved the first result in this direction by explaining how to recover a compact (finite, for instance) group from its finite-dimensional representations. We can now state and prove an analog for affine group schemes.

Let G be an affine group scheme, and let $\mathbf{Rep}_{\mathrm{fin}}^G$ denote the category of finite-dimensional G–representations. There is a forgetful functor

$$\omega\colon \mathbf{Rep}_{\mathrm{fin}}^G \longrightarrow \mathbf{Vec}_{\mathrm{fin}}$$

to the category of finite-dimensional K–vector spaces.

Similarly to what we did before, there is a group $\mathrm{Aut}(\omega)$ of natural isomorphisms $\omega \to \omega$, and a subgroup $\mathrm{Aut}_\otimes(\omega)$ of such isomorphisms $\omega \to \omega$ that are compatible with the tensor product of representations:

$$\sigma(V \otimes W, \Delta_{V\otimes W}) = \sigma(V, \Delta_V) \otimes \sigma(W, \Delta_W)$$

and

$$\sigma(K) = \mathrm{id}_K\,.$$

Here, the second condition is not automatic. Unfortunately, this is not enough to recover G from $\mathrm{Aut}_\otimes(\omega)$; for that, we need to turn $\mathrm{Aut}_\otimes(\omega)$ into an affine group scheme, the same kind of object as G is! We define $\mathrm{Aut}_\otimes(\omega)(B)$ as the group of isomorphisms $\omega \otimes B \to \omega \otimes B$ of functors

$$\mathbf{Rep}_{\mathrm{fin}}^G \longrightarrow \mathbf{Mod}_B,$$

where the target is the category of finitely-generated free B–modules. The elements σ are families of B–linear isomorphisms

$$\sigma(V, \Delta_V)\colon V \otimes B \longrightarrow V \otimes B$$

that are natural with respect to K–linear G–homomorphisms $f\colon V \to W$:

$$\begin{array}{ccc} V \otimes B & \xrightarrow{\sigma(V,\Delta_V)} & V \otimes B \\ {\scriptstyle f\otimes B}\downarrow & & \downarrow{\scriptstyle f\otimes B} \\ W \otimes B & \xrightarrow[\sigma(W,\Delta_W)]{} & W \otimes B. \end{array}$$

Compatibility with the tensor product means that

$$\sigma(V \otimes W, \Delta_{V\otimes W}) = \sigma(V, \Delta_V) \otimes_B \sigma(W, \Delta_W)$$

under the identification $(V \otimes B)_B(W \otimes B) \cong (V \otimes W) \otimes B$, and

$$\sigma(K) = \mathrm{id}_B\,.$$

We are now ready to state the main result.

Theorem 25.3. *There is a canonical isomorphism*

$$G \longrightarrow \mathrm{Aut}_{\otimes}(\omega)$$

of affine group schemes.

Proof. Let us begin by describing the morphism. A B–point $\beta \in G(B)$ is a morphism $\beta\colon H \to B$ of commutative K–algebras. We need to produce a B–linear isomorphism $\varphi_\beta\colon \omega \otimes B \to \omega \otimes B$. That is for each representation (V, Δ_V) a natural isomorphism $\sigma(V, \mathrm{id}_V)\colon V \otimes B \to V \otimes B$. We define this to the unique B–linear extension of $(\mathrm{id}_V \otimes \beta)\Delta_V$. In other symbols, we have a commutative diagram:

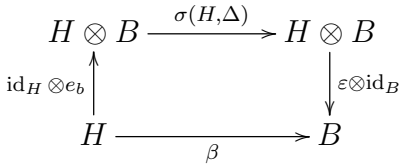
$$\begin{array}{ccc} V & \xrightarrow{\Delta_V} & V \otimes H \\ {\scriptstyle \mathrm{id}_V \otimes e_B}\downarrow & & \downarrow{\scriptstyle \mathrm{id}_V \otimes \beta} \\ V \otimes B & \xrightarrow[\sigma(V,\Delta_V)]{} & V \otimes B \end{array}$$

Let us first show that $\beta \mapsto \varphi_\beta$ is injective. Before we do so, let us remark that any natural isomorphism σ extends uniquely to all co-modules because all of the are unions of their finite-dimensional sub-co-modules by Proposition 22.3. In particular, this applies to the co-module H. We show that $\varphi_\beta(H, \Delta)$ determines β. Indeed, the definition of φ_β gives us a commutative diagram

$$\begin{array}{ccccc} H & \xrightarrow{\Delta} & H \otimes H & \xrightarrow{\varepsilon\otimes\beta} & \\ {\scriptstyle \mathrm{id}_H \otimes e_B}\downarrow & & \downarrow{\scriptstyle \mathrm{id}_H \otimes \beta} & & \\ H \otimes B & \xrightarrow[\sigma(H,\Delta)]{} & H \otimes B & \xrightarrow[\varepsilon\otimes\mathrm{id}_B]{} & B \end{array}$$

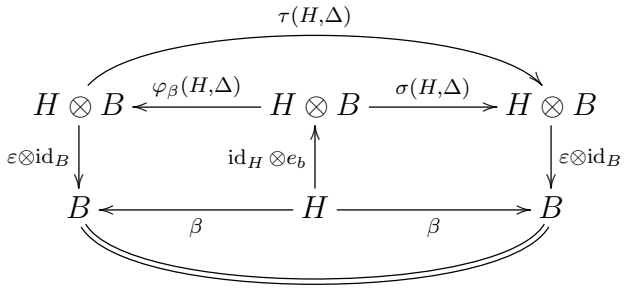
Since $(\varepsilon \otimes \beta)\Delta = \beta$, this proves that β is determined by $\varphi_\beta(H, \Delta)$ and, therefore, by φ_β.

As for surjectivity, we turn the previous argument backward. Given $\sigma \in \mathrm{Aut}_{\otimes}(\omega)$, we define β so that the following diagram commutes



Then it suffices to show the equality $\varphi_\beta = \sigma$ for that β.

Actually, we show that the difference $\tau = \sigma\varphi_\beta^{-1}$ is the identity. For the difference τ, we know that the diagram



commutes, and this means

$$(\varepsilon \otimes \mathrm{id}_B)\tau(H,\Delta) = \varepsilon \otimes \mathrm{id}_B\,. \tag{25.1}$$

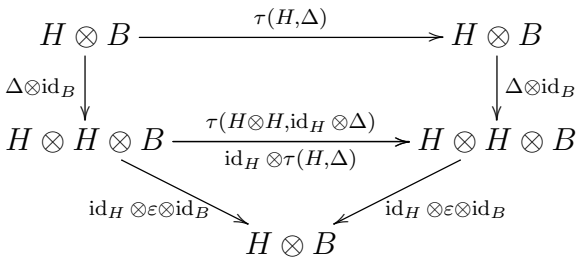
Proposition 24.1 implies that the equality $\tau(V,\Delta_V) = \mathrm{id}_V$ holds for all (V,Δ_V) if and only if it holds for (H,Δ).

We consider $H \otimes H$ with the trivial action on the left factor and the standard (regular) co-action on the right factor. Then the co-multiplication $\Delta\colon H \to H \otimes H$ is a homomorphism of co-modules, and

$$\tau(H \otimes H, \mathrm{id}_H \otimes\Delta) = \mathrm{id}_H \otimes\tau(H,\Delta)$$

by the compatibility of τ with the tensor product. This gives another commutative diagram, where the rectangle is naturality with respect to the homomorphism Δ

of co-modules, and the triangle is (25.1), tensored from the left with id_H:



The outer vertical compositions are both $(\mathrm{id}_H \otimes\varepsilon \otimes \mathrm{id}_B)(\Delta \otimes \mathrm{id}_B) = \mathrm{id}_{H\otimes B}$. Therefore, we can read off that $\tau(H, \Delta) = \mathrm{id}_{H\otimes B}$, finishing the proof. □

Remark 25.4. Tannaka's results are usually complement by the results of Krein who characterized the categories of the form $\mathbf{Rep}^G_{\mathrm{fin}}$ with a (long) list of axioms.

Appendices

A Category language

There is no need to know category *theory* to learn Lie theory. But, it is very helpful to be able to speak, or at least understand, the category *language* because this makes the formulation of some of the results easier. In this appendix, we collect the minimum vocabulary. There are plenty of examples to illustrate these concepts throughout the text.

Definition A.1. A *category* $\mathbf{C}$ is given by rather a lot of data: (1) a 'collection' of *objects* $x, y, \ldots$, (2) for each pair of objects a set $\mathrm{Mor}_{\mathbf{C}}(x,y) = \mathbf{C}(x,y)$ of morphisms $a\colon x \to y$ with *source* x and *target* y, (3) for each object x an *identity* $\mathrm{id}_x\colon x \to x$, and (4) for each pair $a\colon x \to y$ and $b\colon y \to z$ of composable morphism a morphism $b \circ a\colon x \to z$, their *composition*. These data are required to satisfy associativity and neutrality in the obvious sense, which is $\mathrm{id}_y \circ a = a = a \circ \mathrm{id}_a$ and $(a \circ b) \circ c = a \circ (b \circ c)$.

Remark A.2. We use the word 'collection' instead of 'set' because we do not want to worry about set theory and its size restrictions here.

A morphisms $a\colon x \to y$ in a category is called an *isomorphism* if there is a morphism $b\colon y \to x$ such that $a \circ b = \mathrm{id}_y$ and $b \circ a = \mathrm{id}_x$.

Definition A.3. If $\mathbf{C}$ and $\mathbf{D}$ are two categories, a *functor* $F\colon \mathbf{C} \to \mathbf{D}$ assigns an object $F(x)$ of $\mathbf{D}$ to each object x of $\mathbf{C}$ and a morphism $F(a)\colon F(x) \to F(y)$ to each morphism $a\colon x \to y$ in $\mathbf{C}$. Note that the notation already indicates that F has to be compatible with source and target, but it also has to be compatible with identities and composition: we require $F(\mathrm{id}_x) = \mathrm{id}_{F(x)}$ and $F(b \circ a) = F(b) \circ F(a)$.

Example A.4. For every category $\mathbf{C}$ and every object x of $\mathbf{C}$ there is a functor

$$\mathbf{C}(x, ?)\colon \mathbf{C} \longrightarrow \mathbf{Set},\ y \longmapsto \mathbf{C}(x,y)$$

to the category $\mathbf{Set}$ of sets. This is the functor *represented* by the object x.

Definition A.5. Let $F, G\colon \mathbf{C} \to \mathbf{D}$ be two functors defined between the same categories $\mathbf{C}$ and $\mathbf{D}$. A *natural transformation* $\Phi\colon F \to G$ is given by a 'collection' of morphisms $\Phi(x)\colon F(x) \to G(x)$, one for each object x in $\mathbf{C}$, such that for all morphisms $a\colon x \to y$ in $\mathbf{C}$, the diagram

$$\begin{array}{ccc} F(x) & \xrightarrow{\Phi(x)} & G(x) \\ {\scriptstyle F(a)}\downarrow & & \downarrow{\scriptstyle G(a)} \\ F(y) & \xrightarrow[\Phi(y)]{} & G(y) \end{array}$$

commutes. A natural transformation is called a *natural isomorphism* if all morphisms $\Phi(x)$ are isomorphisms.

Remark A.6. In this text, the word 'natural' *always* refers to a natural transformation that is explicit or implicit in the situation. In contrast, the word 'canonical' has no mathematical meaning; it can usually be replaced by 'the first we can think of.'

Exercise 70. Let $\mathbf{C}$ be a category and let $F\colon \mathbf{C} \to \mathbf{Set}$ be a functor to the category of sets. Show that evaluating natural transformations $\mathbf{C}(x, ?) \to F$ at the identify id_x gives a bijection between all such transformations and the set $F(x)$. This observation is called the *Yoneda Lemma*.

A product of two objects y and z in a category $\mathbf{C}$ is an object $y \times z$ together with a natural bijection

$$\mathbf{C}(x, y) \times \mathbf{C}(x, z) \cong \mathbf{C}(x, y \times z).$$

We use this bijection to identify a morphism $x \to y \times z$ with a pair (a, b) of morphisms $a\colon x \to y$ and $b\colon x \to z$. The identity $\mathrm{id}_{y\times z}$ corresponds to the pair $(\mathrm{pr}_y, \mathrm{pr}_z)$ of the projections.

Definition A.7. An *adjunction* consists of two functors $L\colon \mathbf{C} \to \mathbf{D}$ and $R\colon \mathbf{D} \to \mathbf{C}$ together with natural bijections

$$\mathbf{D}(L(x), y) \cong \mathbf{C}(x, R(y)),$$

where x is on object of $\mathbf{C}$ and y is an object of $\mathbf{D}$.

Exercise 71. Find the category with objects (x, y), for x an object of $\mathbf{C}$ and y an object of $\mathbf{D}$, such that $\mathbf{D}(L(x), y)$ and $\mathbf{C}(x, R(y))$ are both functors from that category to the category of sets. This is necessary to make sense of the word natural in Definition A.7.

B Semi-direct products

In this appendix, we shall review some facts about direct and semi-direct products of groups that are not as widely known as they should be. Most of the proofs are straightforward and left as exercises.

B.1 Direct products

The easiest way to decompose a group H is to write it as a direct product

$$H = P \times Q.$$

This is just the product in the category of groups, so that we have a natural bijection

$$\mathrm{Mor}(G, P \times Q) \cong \mathrm{Mor}(G, P) \times \mathrm{Mor}(G, Q)$$

of groups, and this describes all morphisms of groups *into* a direct product.

Can we also say what morphisms *out of* a direct product are? Yes, we can:

The groups P and Q can be embedded into the product $P \times Q$ using $p \mapsto (p, e_Q)$ and $q \mapsto (e_P, q)$. Because of

$$(p, q) = (p \cdot e_P, e_Q \cdot q) = (p, e_Q) \cdot (e_P, q),$$

the product is generated by the images of these embeddings, so that a morphism out of the direct product is also determined by a pair of morphisms out of the factors.

Proposition B.1. *There is a natural bijection*

$$\mathrm{Mor}(P \times Q, G) \cong \{\, (\alpha, \beta) \in \mathrm{Mor}(G, P) \times \mathrm{Mor}(G, Q) \mid [\,\mathrm{Im}(\alpha), \mathrm{Im}(\beta)\,] = e_G \,\}.$$

Exercise 72. Finish the proof by verifying that, (1) given ψ, the images of

$$\alpha \colon p \mapsto \psi(p, e_Q)$$
$$\beta \colon q \mapsto \psi(e_P, q)$$

commute, and (2) for such α and β, the map

$$\psi\colon (p,q) = \alpha(p)\beta(q)$$

is a morphism of groups.

B.2 Extensions of groups

There are more general ways to decompose groups than as direct products. A rather general context is the following. An *extension* of a group Q by a group K is a sequence

$$K \xrightarrow{j} E \xrightarrow{q} Q \tag{B.1}$$

such that j is an embedding onto the kernel of q, and q identifies Q with the quotient $E/j(K)$.

Example B.2. A product $P \times Q$ can be written as an extension of Q by P. It can also be written as an extension of P by Q.

Product extensions are sometimes called *trivial*.

Example B.3. The extension

$$\mathbb{Z}/2 \longrightarrow \mathbb{Z}/4 \longrightarrow \mathbb{Z}/2 \tag{B.2}$$

is not trivial.

Beyond trivial extensions, there is another class of extensions that is fairly easy to understand: An extension is *splittable* if the morphism q has a section. This means that there is a morphism $s\colon Q \to E$ with $qs = \mathrm{id}_Q$.

Example B.4. Product extensions are splittable.

Example B.5. The extension (B.2) is not splittable.

B.3 Semi-direct products

Given a pair of groups K and Q and a morphism

$$\varphi\colon Q \longrightarrow \mathrm{Aut}(K),\ q \mapsto \varphi_q$$

of groups, we can define a group structure on the cartesian product $K \times Q$ by

$$(k,q)\cdot(k',q') = (k\cdot\varphi)q(k'), q\cdot q').$$

$$(k,q)\cdot(k',q') = (k\cdot\varphi_q(k'), q\cdot q').$$

Exercise 73. Check that this is associative and that (e_K, e_Q) is a unit. What is the inverse of (k,q)?

The group defined above is the *semi-direct* product

$$K \rtimes_\varphi Q$$

of the groups K and Q with respect to the action of Q on K via φ.

Example B.6. If φ is trivial, then the semi-direct product is just the usual product.

Exercise 74. Show that the group $K \rtimes_\varphi Q$ can be written as a splittable extension of the group Q by the group K.

Exercise 75. Given an extension B.1 with a splitting s, define an action of Q on K by conjugation:

$$\varphi_q(k) = s(q)\cdot k\cdot s(q)^{-1}.$$

Show that

$$K \rtimes_\varphi Q \longrightarrow E,\ (k,q) \longmapsto k\cdot s(q)$$

is an isomorphism of groups.

Proposition B.7. *There is a natural bijection between* $\mathrm{Mor}(K \rtimes_\varphi Q, G)$ *and*

$$\{\,(\alpha,\beta) \in \mathrm{Mor}(K,G)\times\mathrm{Mor}(Q,G) \mid \alpha\cdot\varphi_q = \beta(q)\cdot\alpha\cdot\beta(q)^{-1} \textit{ for all } q\,\}.$$

The equation $\alpha \cdot \varphi_q = \beta(q) \cdot \alpha \cdot \beta(q)^{-1}$ means that the diagram

$$\begin{array}{ccc} K & \xrightarrow{\varphi_q} & K \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\alpha} \\ G & \xrightarrow[\beta(q)\triangleright]{} & G \end{array}$$

commutes for all q.

Exercise 76. Proof the preceding proposition.

Proposition B.8. *There is a natural bijection between* $\mathrm{Mor}(G, K \rtimes_\varphi Q)$ *and*

$$\{\, (\alpha, \beta) \in \mathrm{Map}(G, K) \times \mathrm{Mor}(G, Q) \mid \alpha(gh) = \alpha(g) \cdot \varphi_{\beta(g)}\alpha(h) \textit{ for all } g, h \,\}.$$

Note that the first factor α is just a map of sets, in general.

Exercise 77. Proof the preceding proposition.

Exercise 78. Try to define a category of split extensions, a category of semi-direct products, and establish an equivalence of categories between them. This is easier when the quotient group Q is fixed throughout.

C Multilinear algebra

This is not a textbook on linear algebra or homological algebra, but we will use tensor products, and we might just as well explain their main features from the 'right' perspective.

C.1 Tensor products of free modules

Let us work with modules over a commutative ring R. For us, the most important cases are when R is a field or the ring of integers.

Let M and N be R–modules.

When M and N are free with bases B and C, their tensor product $M \otimes N$ is free with basis $B \times C$. The basis element (b, c) is written $b \otimes c$, and a general tensor $m \otimes n$ can be expanded bilinearly in terms of these.

Exercise 79. Not every element of the tensor product $M \otimes N$ can be written as a tensor $m \otimes n$. For instance, if $R = K$ is a field, and $M = Kb_1 \oplus Kb_2$, and $N = Kc_1 \oplus Kc_2$ are both 2–dimensional with the indicated bases, then

$$\lambda_{1,1}\, b_1 \otimes c_1 + \lambda_{1,2}\, b_1 \otimes c_2 + \lambda_{2,1}\, b_2 \otimes c_1 + \lambda_{2,2}\, b_2 \otimes c_2$$

can be written in the form $m \otimes n$ if and only if $\lambda_{1,1}\lambda_{2,2} = \lambda_{1,2}\lambda_{2,1}$.

If $R = K$ is a field, then this is all there is to the tensor product. In general, we will have to deal with R–modules that are not free. The situation is still easy when one of the factors is free: If $M = \bigoplus_{b \in B} R$ is still free but N is maybe not, we can still compute

$$M \otimes N = \bigoplus_{b \in B} N,$$

and similarly if N is free and M may not be.

C.2 Tensor products using free resolutions

In general, we will have to 'resolve' M or N or both by free modules. For $R = \mathbb{Z}$ this means that we write M as a quotient F_0/F_1 of a free abelian group F_0 by a subgroup F_1, which will automatically be free. (This is false over more general commutative rings.) For instance, we can take F_0 to be the free module with basis the elements $m \in M$, and the obvious surjective homomorphism $F_0 \to M$ that sends m to m. This is good for theory, but in practice, smaller options are usually available. The cyclic group $\mathbb{Z}/n$ can be resolved by $F_0 = \mathbb{Z}$, for example.

Using tensor products of free abelian groups, we can then form the tensor product

$$F_0 \otimes G_0 \longleftarrow F_1 \otimes G_0 \oplus F_0 \otimes G_1 \longleftarrow F_1 \otimes G_1,$$

where the homomorphisms are induced by the inclusions, except for a sign that is introduced somewhere to make the composition zero. In turns out, still assuming $R = \mathbb{Z}$, that the map to the right is always injective, and we get a homomorphism

$$F_0 \otimes G_0 \longleftarrow \frac{F_1 \otimes G_0 \oplus F_0 \otimes G_1}{F_1 \otimes G_1}.$$

The cokernel is $M \otimes N$, and the kernel is $\operatorname{Tor}(M, N)$, by definition. It would be more prefcise to write $M \otimes_{\mathbb{Z}} N$ and $\operatorname{Tor}_1^{\mathbb{Z}}(M, N)$, respectively.

Example C.1. Let us work out the tensor product $\mathbb{Z}/m \otimes \mathbb{Z}/n$ of two cyclic groups of order $m, n \geqslant 2$. The easiest way to do that might be to resolve the cyclic group $\mathbb{Z}/m$ by $m\colon \mathbb{Z} \to \mathbb{Z}$, and then to tensor this with the other factor $\mathbb{Z}/n$ to get

$$\mathbb{Z}/n \xrightarrow{m} \mathbb{Z}/n.$$

The image consists of the multiples of $\gcd(m, n)$ in $\mathbb{Z}/n$, and the kernel is generated by the quotient $n/\gcd(m, n)$. We get

$$\mathbb{Z}/m \otimes \mathbb{Z}/n = \mathbb{Z}/\gcd(m, n) = \operatorname{Tor}(\mathbb{Z}/m, \mathbb{Z}/n). \tag{C.1}$$

If, for some reason, we decided to resolve both $\mathbb{Z}/m$ and $\mathbb{Z}/n$, we would get the complex

$$\mathbb{Z} \xleftarrow{\begin{bmatrix} m & n \end{bmatrix}} \mathbb{Z} \oplus \mathbb{Z} \xleftarrow{\begin{bmatrix} -n \\ m \end{bmatrix}} \mathbb{Z}$$

The image of the homomorphism on the left is generated by the gcd of m and n. The kernel consists of all

$$\begin{bmatrix} a \\ b \end{bmatrix}$$

such that $am + bn = 0$. As the diagram indicates, the pair

$$\begin{bmatrix} -n \\ m \end{bmatrix}$$

is an obvious choice, but the kernel actually consists of the multiples of

$$\begin{bmatrix} -n/\gcd(m,n) \\ m/\gcd(m,n) \end{bmatrix},$$

and we end up with the same result (C.1).

C.3 The tensor algebra

Let V be a R–module. For each integer $n \geqslant 0$ we can form the n–th tensor power

$$\mathrm{T}^n(V) = \underbrace{V \otimes V \otimes \cdots \otimes V}_{n}.$$

These form a graded object $(\,\mathrm{T}^n(V) \,|\, n \in \mathbb{Z}\,)$, with the convention $\mathrm{T}^n(V) = 0$ if $n < 0$. We can also form the sum

$$\mathrm{T}(V) = \bigoplus_{n=0}^{\infty} \mathrm{T}^n(V),$$

where V sits in degree 1 as $\mathrm{T}^1(V) = V$. The R–module $\mathrm{T}(V)$ becomes an associative algebra with unit $1 \in R = \mathrm{T}^0(V)$ and 'tautological' product

$$(x_1 \otimes x_2 \otimes \cdots \otimes x_m) \otimes (y_1 \otimes y_2 \otimes \cdots \otimes y_n) = x_1 \otimes x_2 \otimes \cdots \otimes x_m \otimes y_1 \otimes y_2 \otimes \cdots \otimes y_n.$$

This is the *tensor algebra* of V. It has the following universal property.

Proposition C.2. *There are natural bijections*

$$\mathbf{Ass}(\mathrm{T}(V), A) \cong \mathbf{Mod}(V, A),$$

whenever V is a R–module and A is an associative R–algebra.

D The Möbius function

The Appendix to [Bou72, Ch. 2] contains a minimal presentation. We write

$$\mathbb{N}^\times = \{1, 2, 3, \dots\}$$

for the multiplicative monoid of positive integers. We define the Möbius function

$$\mu\colon \mathbb{N}^\times \longrightarrow \mathbb{Z}$$

by setting $\mu(n) = 0$ if n is divisible by the square of a prime, and $\mu(n) = (-1)^k$ if k is the number of distinct prime factors of n otherwise.

Proposition D.1. *The Möbius function is the unique function such that* $\mu(1) = 1$ *and*

$$\sum_{d|n} \mu(d) = 0 \tag{D.1}$$

for all $n \geqslant 2$.

Proof. The uniqueness is clear because the properties can be used to compute $\mu(n)$ inductively for any n. It remains to be shown that μ satisfies (D.1).

Let $\mathbb{P}$ denote the infinite set of primes and $\mathbb{P}(n) \subseteq \mathbb{P}$ the subset of prime numbers that divide n. The prime factor decomposition of n takes the form

$$n = \prod_{p \in \mathbb{P}(n)} p^{v_p(n)}$$

with exponents $v_p(n) \geqslant 1$. We have $v_p(n) = 0$ if and only $p \notin \mathbb{P}(n)$. Since μ vanishes on the numbers divisible by squares of primes, we can concentrate on the square-free divisors of n. The square-free divisors of n correspond to the subsets $S \subseteq \mathbb{P}(n)$, and the value of the Möbius function on the divisor corresponding to the subset S is $(-1)^{|S|}$:

$$\sum_{d|n} \mu(d) = \sum_{S \subseteq \mathbb{P}(n)} (-1)^{|S|}.$$

We now reorder this by cardinality to get

$$\sum_{s=0}^{|\mathbb{P}(n)|} \binom{|\mathbb{P}(n)|}{s} (-1)^s = (1-1)^{|\mathbb{P}(n)|} = 0$$

because $|\mathbb{P}(n)| \geqslant 1$ if $n \geqslant 2$. □

Theorem D.2. (Möbius inversion) *Let $f, g \colon \mathbb{N}^\times \to A$ be two functions with values in an abelian group, written additively. Then the equation*

$$f(n) = \sum_{d|n} g(d) \tag{D.2}$$

for all $n \geqslant 1$ is equivalent to the equation

$$g(n) = \sum_{d|n} \mu(d) f(n/d) \tag{D.3}$$

for all $n \geqslant 1$.

Proof. Let us first assume (D.2) and show (D.3). We have

$$\sum_{d|n} \mu(d) f(n/d) = \sum_{d|n} \mu(d) \sum_{c|\frac{n}{d}} g(c)$$

by (D.2). We can write this as

$$\sum_{cd|n} \mu(d) g(c),$$

where the sum is over all pairs (c, d) such that $cd|n$. The same argument shows that this equals

$$\sum_{c|n} g(c) \sum_{d|\frac{n}{c}} \mu(d) = g(n)$$

because $\sum_{d|\frac{n}{c}} \mu(d) = 0$ unless $c = n$, by (D.1).

Let us last assume (D.3) and show (D.2). We have

$$\sum_{d|n} g(d) = \sum_{d|n} \sum_{b|d} \mu(b) f(d/b).$$

We write this as

$$\sum_{b|d|n} \mu(b) f(d/b),$$

the sum being over all (b, d) such that $b|d|n$. Writing $d = ab$, this becomes

$$\sum_{a|nb} \mu(b) f(a) = \sum_{a|n} f(a) \sum_{b|\frac{n}{a}} \mu(b) = f(n)$$

because $\sum_{b|\frac{n}{a}} \mu(b) = 0$ unless $a = n$, by (D.1). □

Example D.3. Let $\Phi_n(\mathrm{T})$ be the n–th cyclotomic polynomial,

$$\Phi_n(\mathrm{T}) = \prod_z (\mathrm{T} - z),$$

where the product is over the set of *primitive* n–th roots of unity or, equivalently, the generators of the group of n–th roots of unity. Then

$$\mathrm{T}^n - 1 = \prod_{d|n} \Phi_d(\mathrm{T}),$$

and Möbius inversion gives

$$\Phi_n(\mathrm{T}) = \prod_{d|n} (\mathrm{T}^{n/d} - 1)^{\mu(d)}.$$

Exercise 80. Show that the degree of $\Phi_n(\mathrm{T})$ is

$$n \prod_{p \in \mathbb{P}(n)} (1 - p^{-1}).$$

Let D be the ring of *formal Dirichlet series*. The elements of the ring D are all functions $f \colon \mathbb{N}^\times \to \mathbb{C}$. They are added point-wise, and multiplied using convolution:

$$(f \cdot g)(n) = \sum_{d|n} f(d) g(n/d).$$

Note that this is well-defined, even if f and g are not finitely supported. The convolution product imitates the multiplication of the series $\sum_{n=1}^{\infty} f(n) n^{-s}$, without worrying about convergence.

Exercise 81. Show that f is invertible if and only if $f(1) \neq 0$.

Exercise 82. Let μ be the Möbius function, and let ζ be the constant function 1. Show that μ and ζ are inverse to each other. Paraphrasing Möbius inversion, the equations $f = \zeta \cdot g$ and $g = \mu \cdot f$ are equivalent to each other.

Exercise 83. Let n^{-s} denote the characteristic function of n. Show that the product

$$\prod_{p \in \mathbb{P}} (1 - p^{-s})$$

makes sense and equals μ. This Euler product reflects the unique prime decomposition in $\mathbb{Z}$. Similarly,

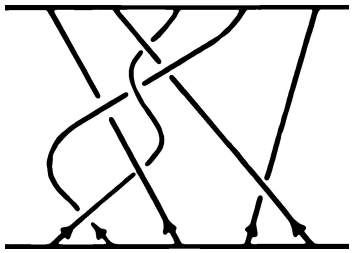
$$\zeta = \prod_{p \in \mathbb{P}} (1 - p^{-s})^{-1}.$$

E Student projects

This appendix collects a couple of topics that are suitable for projects after taking this course. I only listed some that are close to my interests. Let me know if you would like me to find something that is closer to yours.

E.1 Braid groups and the Drinfeld–Kohno theorem

Braids, in the mathematical sense, describe how n points in the plane can be connected to n points in a parallel plane without the paths crossing each other or going backwards in between.



For a given n, the braids from a group B_n, with composition given by stacking braids on top of each other, inversion by vertical reflection, and identity element the straight braid. There is a surjective morphism $\mathrm{B}_n \to \mathrm{S}_n$ to the symmetric group S_n which keeps track of which points are connected to which points. The kernel is the pure braid group P_n. An easy geometric argument shows that we can write P_n as a semi-direct product (see Appendix B): $\mathrm{P}_n = \mathrm{F}_{n-1} \rtimes \mathrm{P}_{n-1}$, so that, by induction, the pure braid group is an iterated semi-direct product of free groups. Just as we did for free groups, it is possible to describe the Lie algebra $\mathrm{Gr}(\mathrm{P}_n)$. This leads to the Drinfeld–Kohno Lie algebra and the Knizhnik–Zamolodchikov differential equations.

E.2 Lie algebras in homotopy theory

Recall that we can construct the torus $\mathrm{S}^1 \times \mathrm{S}^1$ by attaching a 2–cell to the figure eight $\mathrm{S}^1 \vee \mathrm{S}^1$ along a map $\mathrm{S}^1 \to \mathrm{S}^1 \vee \mathrm{S}^1$ that represents the commutator $[x, y]$ in the fundamental group of the figure-eight, which is free on x and y. More generally, the attaching map $\mathrm{S}^{m+n-1} \to \mathrm{S}^m \vee \mathrm{S}^n$ of the top-cell of the product $\mathrm{S}^m \times \mathrm{S}^n$ defines a natural operation $\pi_m(X) \otimes \pi_n(X) \to \pi_{m+n-1}(X)$ on the homotopy groups of a space X, and this turns the graded abelian group $\bigoplus_{n \geqslant 2} \pi_n(X)$ into some sort of graded Lie algebra. There are some minor complications, due to the fact that the fundamental group $\pi_1(X)$ need not be abelian, so that it is easier to assume that X is simply-connected, and some signs which have to be worked in from the grading. Also, the homotopy groups usually have torsion, and the product does not satisfy the stronger condition that $[x, x] = 0$. That's one reason why working rationally makes things easier still. One noteworthy theorem in this context is the Hilton–Milnor theorem.

E.3 Morse theory on the unitary groups

Morse theory is a method of studying the topology of a manifold M using particularly nice functions $f \colon M \to \mathbb{R}$: those where the second derivative is non-degenerate in those point where the first derivative vanishes. It is possible to describe Morse functions on the unitary groups

$$\mathrm{U}(n) = \{\, A \in \mathrm{GL}_n(\mathbb{C}) \mid A^*A = E = AA^* \,\}$$

explicitly, and use them to deduce topological information about these groups, using linear algebra. Once this is understood, there are plenty of opportunities for generalization.

E.4 Quantum groups and Yang–Baxter equations

Quantum groups are not groups, but Hopf algebras. They are Hopf algebras that are 'deformations' of universal enveloping algebras of Lie algebras or of Hopf algebras of functions on algebraic groups, but they are neither commutative nor cocommutative. Their interest lies mainly in their categories of representations V, which have symmetries not present in categories of representations of ordinary groups or Lie algebras. These symmetries $S\colon V \otimes V \to V \otimes V$ give rise to nontrivial solutions of the (parameter-independent) Yang–Baxter equation

$$(S \otimes \mathrm{id})(\mathrm{id} \otimes S)(S \otimes \mathrm{id}) = (\mathrm{id} \otimes S)(S \otimes \mathrm{id})(\mathrm{id} \otimes S)$$

in statistical mechanics and quantum field theory.

E.5 Moduli spaces of Lie algebras

Given a finite-dimensional Lie algebra L, we can choose a basis $x_1, \dots, x_n$ and write the bracket in the form

$$[x_i, x_j] = \sum_{k=1}^{n} \lambda_{i,j}^k x_k$$

with some structure constants $\lambda_{i,j}^k$ in the ground field. Which tuples $\lambda_{i,j}^k$ arise this way? What are the equations that the structure constants have to satisfy? How does the description transform under a change of basis? What are the symmetry groups here? More conceptually, an alternating multiplication on L is given by a linear map $\Lambda^2 L \to L$, or by an element in the vector space $\mathrm{Hom}(\Lambda^2 L, L)$. The Jacobi cuts out the locus of points that correspond to Lie algebras. What does it look like? The general linear group $\mathrm{GL}(L)$ preserves this locus, and the orbit space (in a suitable sense) is a moduli space of Lie algebras. There are plenty of opportunities to explore this with low-dimensional computations and variants for other kinds of algebras.

E.6 Torus actions and the geometry of Gröbner bases

Bibliography

[AB12] R. Achilles, A. Bonfiglioli. The early proofs of the theorem of Campbell, Baker, Hausdorff, and Dynkin. Arch. Hist. Exact Sci. 66 (2012) 295–358.

[Ada69] J.F. Adams. Lectures on Lie Groups. W.A. Benjamin, Inc., 1969.

[Ale59] P.S. Alexandroff. An Introduction to the Theory of Groups. Hafner Publishing Co., Inc., 1959.

[AF09] N. Andruskiewitsch, W. Ferrer Santos. The beginnings of the theory of Hopf algebras. Acta Appl. Math. 108 (2009) 3–17.

[Bak02] A. Baker. Matrix Groups. An Introduction to Lie Group Theory. Springer Undergraduate Mathematics Series. Springer-Verlag, 2002.

[Ber78] G.M. Bergman. The diamond lemma for ring theory. Adv. in Math. 29 (1978) 178–218.

[Ber85] G.M. Bergman. Everybody knows what a Hopf algebra is. Group actions on rings (Brunswick, Maine, 1984) 25–48. Contemp. Math. 43. Amer. Math. Soc., Providence, RI, 1985.

[Ber15] G.M. Bergman. An Invitation to General Algebra and Universal Constructions. Universitext. Springer, 2015.

[Beu92] F. Beukers. Differential Galois theory. From number theory to physics (Les Houches, 1989) 413–439. Springer, Berlin, 1992.

[Bir37] G. Birkhoff. Representability of Lie algebras and Lie groups by matrices. Ann. of Math. 38 (1937) 526–532.

[BR89] G. Birkhoff, G.-C. Rota. Ordinary differential equations. Fourth edition. John Wiley & Sons, Inc., New York, 1989.

[BF12] A. Bonfiglioli, R. Fulci. Topics in noncommutative algebra. The theorem of Campbell, Baker, Hausdorff and Dynkin. Lecture Notes in Mathematics 2034. Springer, Heidelberg, 2012.

[Bor01] A. Borel. Essays in the History of Lie Groups and Algebraic Groups. History of Mathematics 21. American Mathematical Society, Providence, RI, 2001.

[Bou60] N. Bourbaki. Groupes et algèbres de Lie. Chap. 1: Algèbres de Lie. Actualites scientifiques et industrielles 1285. Paris, Hermann & Cie, 1960.

[Bou68] N. Bourbaki. Groupes et algèbres de Lie. Chapitres IV, V et VI: Groupes de Coxeter et systèmes de Tits. Groupes engendrés par des réflexions. Systèmes de racines. Actualités Scientifiques et Industrielles 1337. Paris, Hermann & Cie, 1968.

[Bou72] N. Bourbaki. Groupes et algèbres de Lie. Chap. II: Algèbres de Lie libres. Chap. III: Groupes de Lie. Actualites scientifiques et industrielles 1349. Paris, Hermann, 1972.

[Bou75] N. Bourbaki. Groupes et algèbres de Lie. Chap. VII: Sous-algèbres de Cartan, éléments réguliers. Chap. VIII: Algèbres de Lie semi-simples déployées. Actualités Scientifiques et Industrielles 1364. Paris, Hermann, 1975.

[Bou82] N. Bourbaki. Groupes et algèbres de Lie, Chapitre IX: Groupes de Lie réels compacts. Paris, Masson, 1982.

[BtD85] T. Bröcker, T. tom Dieck. Representations of Compact Lie Groups. Graduate Texts in Mathematics 98. Springer-Verlag, New York, 1985.

[CE56] H. Cartan, S. Eilenberg. Homological Algebra. Princeton University Press, Princeton, N. J., 1956.

[CSM95] R. Carter, G. Segal, I. Macdonald. Lectures on Lie Groups and Lie Algebras. London Mathematical Society Student Texts 32. Cambridge University Press, Cambridge, 1995.

[Car58] P. Cartier. Remarques sur le théorème de Birkhoff–Witt. Ann. Scuola Norm. Sup. Pisa Cl. Sci. 12 (1958) 1–4.

[Car07] P. Cartier. A primer of Hopf algebras. Frontiers in Number Theory, Physics, and Geometry. II, 537–615. Springer, Berlin, 2007.

[Car10] P. Cartier. Vinberg algebras, Lie groups and combinatorics. Quanta of maths, 107–126. Clay Math. Proc. 11. Amer. Math. Soc., Providence, RI, 2010.

[CL01] F. Chapoton, M. Livernet. Pre-Lie algebras and the rooted trees operad. Internat. Math. Res. Notices 8 (2001) 395–408.

[Coh17] B. Cohen. A generalization of the Hall–Witt identity. Israel J. Math. 221 (2017) 605–636.

[Coh63] P.M. Cohn. A remark on the Birkhoff–Witt theorem. J. London Math. Soc. 38 (1963) 197–203.

[CH11] T. Crespo, Z. Hajto. Algebraic groups and differential Galois theory. Graduate Studies in Mathematics 122. American Mathematical Society, Providence, RI, 2011.

[DG70] M. Demazure, P. Gabriel. Groupes algèbriques. Tome I: Géométrie algébrique, généralités, groupes commutatifs. Masson & Cie, Éditeur, Paris; North-Holland Publishing Co., Amsterdam, 1970.

[Eic68] M. Eichler. A new proof of the Baker–Campbell–Hausdorff formula. J. Math. Soc. Japan 20 (1968) 23–25.

[FM-K18] G. Fløystad, H. Munthe-Kaas. Pre- and post-Lie algebras: the algebro-geometric view. Computation and combinatorics in dynamics, stochastics and control, 321–367. Abel Symp. 13. Springer, Cham, 2018.

[Gar10] S. Garibaldi. What is...a linear algebraic group? Notices Amer. Math. Soc. 57 (2010) 1125–1126.

[Gri04] P.-P. Grivel. Une histoire du théorème de Poincaré–Birkhoff–Witt. Expo. Math. 22 (2004) 145–184.

[Hig69] P.J. Higgins. Baer invariants and the Birkhoff–Witt theorem. J. Algebra 11 (1969) 469–482.

[HL11] J.H. Hubbard, B.E. Lundell. A first look at differential algebra. Amer. Math. Monthly 118 (2011) 245–261.

[How83] R. Howe. Very basic Lie theory. Amer. Math. Monthly 90 (1983) 600–623. Correction to: "Very basic Lie theory". Amer. Math. Monthly 91 (1984) 247.

[Iwa48] K. Iwasawa. On the representation of Lie algebras. Jpn. J. Math. 19 (1948) 405–426.

[Kap57] I. Kaplansky. An introduction to differential algebra. Actualités Sci. Ind. 1251. Hermann, Paris, 1957.

[Kas95] C. Kassel. Quantum Groups. Graduate Texts in Mathematics 155. Springer-Verlag, 1995.

[Kol68] E.R. Kolchin. Algebraic groups and algebraic dependence. Amer. J. Math. 90 (1968) 1151–1164.

[Kug93] M. Kuga. Galois' dream: group theory and differential equations. Translated from the 1968 Japanese original. Birkhäuser, Boston, MA, 1993.

[Laz54] M. Lazard. Sur les algèbres enveloppantes universelles de certaines algèbres de Lie. Publ. Sci. Univ. Alger. Sér. 1 (1954) 281–294.

[Lev90] A.H.M. Levelt. Differential Galois theory and tensor products. Indag. Math. 1 (1990) 439–449.

[Lod87] J.-L. Loday. Comparaison des homologies du groupe linéaire et de son algèbre de Lie. Ann. Inst. Fourier (Grenoble) 37 (1987) 167–190.

[Lod92] J.-L. Loday. Cyclic homology. Grundlehren der Mathematischen Wissenschaften 301. Springer-Verlag, 1992.

[Lod93] J.-L. Loday. Une version non commutative des algèbres de Lie: les algèbres de Leibniz. Enseign. Math. 39 (1993) 269–293.

[Lod03] J.-L. Loday. Algebraic K-theory and the conjectural Leibniz K-theory. K-Theory 30 (2003) 105–127.

[Mac71] S. Mac Lane. Categories for the working mathematician. Graduate Texts in Mathematics 5. Springer-Verlag, 1971.

[Mil84] J. Milnor. Remarks on infinite-dimensional Lie groups. Relativity, Groups and Topology, II (Les Houches, 1983) 1007–1057. North-Holland, Amsterdam, 1984.

[RU07] D. Riley, H. Usefi. The isomorphism problem for universal enveloping algebras of Lie algebras. Algebr. Represent. Theory 10 (2007) 517–532.

[Sau16] J. Sauloy. Differential Galois theory through Riemann–Hilbert correspondence. Graduate Studies in Mathematics 177. American Mathematical Society, Providence, RI, 2016.

[Sch82] W. Schmid. Poincaré and Lie groups. Bull. Amer. Math. Soc. 6 (1982) 175–186.

[Ser65] J.-P. Serre. Lie Algebras and Lie Groups. Lectures given at Harvard University 1964. W.A. Benjamin, Inc., 1965.

[Ser66] J.-P. Serre. Algébres de Lie Semi-Simples Complexes. W.A. Benjamin, inc., 1966.

[Ser93] J.-P. Serre. Gèbres. Enseign. Math. 39 (1993) 33–85.

[Sta15] R.P. Stanley. Catalan numbers. Cambridge University Press, New York, 2015.

[Sti08] J. Stillwell. Naive Lie theory. Undergraduate Texts in Mathematics. Springer, 2008.

[T-TT99] T. Ton-That, T.-D. Tran. Poincaré's proof of the so-called Birkhoff–Witt theorem. Rev. Histoire Math. 5 (1999) 249–284.

[Tu85] G.Z. Tu. An elementary proof of the Campbell–Hausdorff formula. Sci. Exploration 5 (1985) 103–106.

[Tu04] L.W. Tu. Une courte démonstration de la formule de Campbell–Hausdorff. J. Lie Theory 14 (2004) 501–508.

[Wat79] W.C. Waterhouse. Introduction to affine group schemes. Graduate Texts in Mathematics 66. Springer-Verlag, New York-Berlin, 1979.

[Wey52] H. Weyl. Symmetry. Princeton University Press, Princeton, N.J., 1952.

[Wit37] E. Witt. Treue Darstellung Liescher Ringe. J. Reine Angew. Math. 177 (1937) 152–160.