Free Lie algebras

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

The purpose of this essay is to give an introduction to free Lie algebras and a few of their applications.

My principal references are [Serre:1965], [Reutenauer:1993], and [de Graaf:2000]. My interest in free Lie algebras has been motivated by the well known conjecture that Kac-Moody algebras can be defined by generators and relations analogous to those introduced by Serre for finite-dimensional semi-simple Lie algebras. I have had this idea for a long time, but it was coming across the short note [de Graaf:1999] that acted as catalyst for this (alas! so far unfinished) project.

Fix throughout this essay a commutative ring R. I recall that a Lie algebra over R is an R-module $\mathfrak{g}$ together with a Poisson bracket $[x, y]$ such that

$$\begin{aligned}[x, x] &= 0 \\ [x, [y, z]] + [y, [z, x]] + [z, [x, y]] &= 0\end{aligned}$$

Since $[x + y, x + y] = [x, x] + [x, y] + [y, x] + [y, y]$, the first condition implies that $[x, y] = -[y, x]$. The second condition is called the **Jacobi identity**.

In a later version of this essay, I'll discuss the Baker-Campbell-Hausdorff Theorem (in the form due to Dynkin).

Contents

Throughout, let X be a finite set. Elements of X will be considered as letters in an alphabet.

1. Magmas

If R is a commutative ring, an R-algebra is a free module A over R with a product map

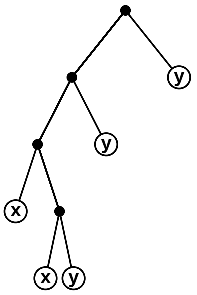
$$A \otimes_R A \longrightarrow A\,.$$

There is no assumption on the product other than bilinearity. Given the finite set X, there is a universal R-algebra generated by it, called a **magma**, described in terms of a basis made up of **monomials**.

A **monomial** is an expression involving X and a pair of brackets $\lfloor, \rfloor$ that are defined recursively: (1) every x in X is a monomial; (2) if p and q are monomials then so is $\lfloor p, q \rfloor$. The magma M_X is the set of all monomials.

The expression $\lfloor p, q \rfloor$ defines a non-associative product in M_X. Every monomial other than an element of X may be factored uniquely as $\lfloor p, q \rfloor$.

These products are essentially **trees**. What this means is that every monomial can be represented by a rooted tree in which leaf nodes are single letters and all other nodes have two monomials as descendants. Here is the tree for $\lfloor \lfloor \lfloor x, \lfloor x, y \rfloor \rfloor, y \rfloor, y \rfloor$:



The **degree** $|m|$ of a monomial m is the number of leaf nodes in the tree, or equivalently the number of elements of X in its expression. Inductively, $|x| = 1$ and $\big| \lfloor p, q \rfloor \big| = |p| + |q|$. The number m_n of monomials of a given degree n is finite, and it is not difficult to evaluate it explicitly.

The monomials of degree one are just the elements of X, so $m_1 = |X|$. There is a simple way to generate inductively all monomials of degree n—each one of these can be expressed uniquely as $\lfloor p, q \rfloor$ with p of degree i and q of degree $n - i$, for $1 \le i \le n - 1$.

$$M_X^{(n)} = \bigsqcup_{1 \le i \le n-1} \big\{ \lfloor p, q \rfloor \,\big|\, p \in M_X^{(i)}, q \in M_X^{(n-i)} \big\} \,.$$

For example, here are the lists of $M_X^{(n)}$ with $X = \{x, y\}$ and $n = 1, 2$:

$$\begin{aligned} & x, y \\ & \lfloor x, x \rfloor, \lfloor x, y \rfloor, \lfloor y, x \rfloor, \lfloor y, y \rfloor \,. \end{aligned}$$

Hence for $n \ge 2$

$$m_n = \sum_{1 \le i \le n-1} m_i m_{n-i}$$

1.1. Proposition. *If $d = |X|$ and*

$$M(t) = m_1 t + m_2 t^2 + \cdots$$

is the generating series for (m_n), then

$$M(t) = \frac{1 - \sqrt{1 - 4dt}}{2} \,.$$

The expression $\sqrt{1 - 4dt}$ here means the series one obtains from the binomial expansion for $(1 + x)^{1/2}$. In calculation this gives an efficient inductive formula for computing $m_n = -c_n/2$ (for $n \ge 1$) with

$$c_0 = 1, \quad c_n = c_{n-1} \cdot d \cdot 2(2n - 3)/n \,.$$

The first few m_n are

$$m_1 = d,\ m_2 = d^2,\ m_3 = 2d^3,\ m_4 = 5d^4,\ m_5 = 14d^5,\ m_6 = 42d^6 \,.$$

Proof. The equation just before the proposition is equivalent to

$$M^2 = M - m_1 t\,.$$

To every monomial is associated its **word**, that is to say the sequence of letters that occur in its leaves, and in order of a left to right traversal. Thus the word of $\lfloor\lfloor x, y\rfloor, \lfloor x, \lfloor x, \lfloor x, y\rfloor\rfloor\rfloor\rfloor$ is $xyxxxy$. Many monomials have the same word.

I'll often label the nodes of a monomial by bit sequences. If h is a compound monomial, h_0 and h_1 will be its left and right components, so that $h = \lfloor h_0, h_1\rfloor$. This continues so that, for example, if h_1 is compound then $h_1 = \lfloor h_{10}, h_{11}\rfloor$ and $h = \lfloor h_0, \lfloor h_{10}, h_{11}\rfloor\rfloor$.

The universal R-algebra $A_{X,R}$ generated by X is the space of linear combinations of monomials in X. This is given the product

$$\Big(\sum_m a_m m\Big) \cdot \Big(\sum_n b_n n\Big) = \sum_{m,n} a_m b_n \lfloor m, n\rfloor\,.$$

Any map from X to an R-algebra A extends to a unique homomorphism from $A_{X,R}$ to A. Usually R will be implicit.

2. The free Lie algebra

For brevity of notation, for any x, y, z in A_X let

$$(\!(x, y, z)\!) = \lfloor x, \lfloor y, z\rfloor\rfloor + \lfloor y, \lfloor z, x\rfloor\rfloor + \lfloor z, \lfloor x, y\rfloor\rfloor\,.$$

Let $\mathcal{L}_X = \mathcal{L}_{X,R}$ be the quotient of $A_X = A_{X,R}$ be the two-sided ideal I generated by terms $\lfloor p, p\rfloor$, $\lfloor p, q\rfloor + \lfloor q, p\rfloor$, and $(\!(p, q, r)\!)$, in which p, q, r are monomials in M_X. Since

$$\Big\lfloor \sum c_m m, \sum c_m m \Big\rfloor = \sum c_m^2 \lfloor m, m\rfloor + \sum_{m \neq n} c_m c_n (\lfloor m, n\rfloor + \lfloor n, m\rfloor)$$

and

$$\Big(\!\!\Big(\sum c_\ell \ell, \sum c_m m, \sum c_n n\Big)\!\!\Big) = \sum_{\ell,m,n} c_\ell c_m c_n (\!(\ell, m, n)\!)$$

the analogous elements of A_X also belong to I. The image $[x, y]$ of $\lfloor x, y\rfloor$ in $\mathcal{L}_X$ depends only on x, y modulo I, and defines the structure of a Lie algebra.

2.1. Lemma. *The ideal of relations is homogeneous and consequently $\mathcal{L}_X$ is graded by degree.*

Proof. Every x in A_X can be expressed as a unique sum of homogenereous components x_n. The claim is that that x lies in I if and only if each x_n lies in I. It suffices to verify that the homogeneous components of each $\lfloor x, x\rfloor$ and each $\lfloor x, \lfloor y, z\rfloor\rfloor + \lfloor y, \lfloor z, x\rfloor\rfloor + \lfloor z, \lfloor x, y\rfloor\rfloor$ lies in I, which is immediate.

Each homogeneous component $\mathcal{L}_X^{(n)}$ is finitely generated over R. We shall see eventually that it is also free over R.

The Lie algebra $\mathcal{L}_X$ has this characteristic property of universality, which is easy to define and verify:

2.2. Proposition. *Any map from X to a Lie algebra $\mathfrak{g}$ extends to a unique Lie algebra homomorphism from $\mathcal{L}_X$ to $\mathfrak{g}$.*

2.3. Corollary. *This property determines the Lie algebra $\mathcal{L}_X$ up to isomorphism.*

Let R_X be the free R-module generated by X. The set X may be identified with a basis of R_X.

If V is any free R-module, for each $n \geq 0$ let $\bigotimes^n V$ be the space of n-tensors of V, and let $\bigotimes^\bullet V$ be the direct sum of the $\bigotimes^n V$. If X is a basis, then the set of products $x_1 \otimes \ldots \otimes x_n$ with each x_i in X is a basis of $\bigotimes^\bullet V$.

If M is any R-module then $\bigwedge^2 M$ is the quotient of $M \otimes M$ by the submodule spanned by the elements $m \otimes m$. It has the universal property that any R-homomorphism from $M \otimes M$ to an R-module that takes any $m \otimes m$ to 0 factors through the projection to $\bigwedge^2 M$.

2.4. Lemma. *If X is assigned a linear order, the R-module $\bigwedge^2 R_X$ has as basis the images $x \wedge y$ of $x \otimes y$ for $x < y$ in X.*

Proof. The elements $x \otimes y$ form a basis of $\bigotimes^2 R_X$, so the $x \wedge y$ certainly span $\bigwedge^2 R_X$. The map taking $x \otimes y$ to $x \otimes y - y \otimes x$ defines a map from $\bigotimes^2 R_X$ to itself. It takes every $m \otimes m$ to 0, hence factors through $\bigwedge^2 R_X$. The image of the $x \wedge y$ is $x \otimes y - y \otimes x$, so any linear relation among these wedge products leads to one of basis elements in $\otimes^2 R_X$, hence must be trivial. ∎

2.5. Proposition. *The canonical map from X to $\mathcal{L}_X$ induces isomorphisms of R_X-modules $R_X \cong \mathcal{L}^1_X$ and $\bigwedge^2 R_X \cong \mathcal{L}^2_X$.*

Proof. The free module R_X is itself an abelian Lie algebra with nil bracket, hence there exists a canonical map from $\mathcal{L}_X$ to R_X, which induces the identity on R_X. This proves the claim about $\mathcal{L}^1_X$. The second claim depends similarly on the Lie algebra which as an R-module is the direct sum $R_X \oplus \bigwedge^2 R_X$ and with brackets

$$\begin{aligned}
[x, y] &= x \wedge y \\
[x, y \wedge z] &= 0 \\
[x \wedge y, z] &= 0 \\
[x \wedge y, z \wedge w] &= 0\,.
\end{aligned}$$ ∎

There is another—and remarkable—avatar of the free Lie algebra $\mathcal{L}_X$, which I'll now explain.

BASE RING EXTENSION. Suppose S to be a commutative ring containing R. If U is any R-module and V an S-module, there is a canonical map

$$\mathrm{Hom}_S(S \otimes_R U, V) \longrightarrow \mathrm{Hom}_R(U, V), \quad f \longmapsto \{u \mapsto f(1 \otimes u)\}\,. \tag{2.6}$$

2.7. Lemma. *In these circumstances, this map is a bijection.*

Proof. The characteristic property of the tensor product is that

$$\mathrm{Hom}_R(S \otimes_R U, V)$$

is the set of R-bilinear maps f from $S \times U$ to V. The subset of elements in $\mathrm{Hom}_S(S \otimes_R U, V)$ are those f such that $f(s \cdot t, u) = s \cdot f(t, u)$ for all s, t in S, u in U. Suppose f in $\mathrm{Hom}_R(U, V)$, and let F be the bilinear map form $S \times U$ to V taking (s, u) to $s \cdot f(u)$. The map $f \mapsto F$ defines an inverse to (2.6) . ∎

2.8. Proposition. *For $R \subseteq S$ the canonical map*

$$S \otimes_R \mathcal{L}_{X,R} \longmapsto \mathcal{L}_{X,S}$$

is an isomorphism.

Proof. The left hand side satisfies the universality property. ∎

RELATIONS WITH THE TENSOR ALGEBRA. Let $\mathfrak{g}$ be an arbitrary Lie algebra over the commutative ring R. Assume that it is free as an R-module. The **universal enveloping algebra** $U(\mathfrak{g})$ of $\mathfrak{g}$ is the quotient of the tensor algebra $\bigotimes^\bullet \mathfrak{g}$ by the two-sided ideal generated by the elements $x \otimes y - y \otimes x - [x, y]$ for x, y in $\mathfrak{g}$. Let $\iota\colon \mathfrak{g} \longmapsto U(\mathfrak{g})$ be the canonical map taking u to its image in the quotient.

The following will be proved in the next section.

2.9. Lemma. *The canonical map ι from $\mathfrak{g}$ to its enveloping algebra is an injection.*

It happens that the universal enveloping algebra of $\mathcal{L}_X$ has a relatively simple realization as a familiar object. Let $\bigotimes^\bullet R_X$ be the tensor algebra generated by R_X. I identify R_X with its copy inside this algebra.

The tensor algebra contains the Lie algebra L_X, defined to be the smallest subspace of $\bigotimes^\bullet R_X$ containing R_X and closed under Poisson bracket $[p,q] = p \otimes q - q \otimes p$. The copies of R_X in $\mathcal{L}_X$ and L_X induce a surjective homomorphism of Lie algebras from $\mathcal{L}_X$ to L_X, and hence a ring homomorphism from the enveloping algebra $U(\mathcal{L}_X)$ to $\bigotimes^\bullet R_X$.

The next results will be proved in the next few sections:

2.10. Theorem. *Each homogeneous component $\mathcal{L}_X^{(n)}$ is a free R-module.*

2.11. Theorem. *The canonical map from $\mathcal{L}_X$ to L_X is an isomorphism.*

2.12. Theorem. *The canonical homomorphism from $U(\mathcal{L}_X)$ to $\bigotimes^\bullet R_X$ is an isomorphism.*

2.13. Theorem. *The rank $\ell_X(n)$ of $\mathcal{L}_X^{(n)}$ depends only on $d = |X|$. It is determined by the recursion*

$$\begin{aligned} \ell_X(1) &= d \\ \ell_X(n) &= \left(\frac{1}{n}\right) \sum_{m|n} \mu(m) d^{n/m} . \end{aligned}$$

Here μ is the classical Möbius function (defined in §16.4 of [Hardy-Wright:1960]):

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1 \\ (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct primes} \\ 0 & \text{otherwise.} \end{cases}$$

The proofs of these results are intertwined, and will be given in the next few sections.

3. Poincaré-Birkhoff-Witt

In this section I digress slightly to include a very brief introduction to universal enveloping algebras. The principal goal is a short account of the elegant proof of the Poincaré-Birkhoff-Witt theorem to be found in [Bergman:1978]. I include this here for two reasons. The first is that it introduces a technique known as 'confluence' that we shall see again. The second is that the PBW theorem will play a crucial role in the remainder of this essay. Bergman's argument has a modern flavour, but is underneath not all that different from the original one of [Birkhoff:1937].

Let $\mathfrak{g}$ be a Lie algebra over R that is free as an R-module. Let X be a basis of $\mathfrak{g}$, and choose a linear order on it. Define a **reduced** monomial $x_1 \otimes \ldots \otimes x_n$ in $\bigotimes^\bullet \mathfrak{g}$ to be one with $x_1 \geq \ldots \geq x_n$. These are part of a basis of $\bigotimes^\bullet \mathfrak{g}$. An **irreducible** tensor is one all of whose monomials are reduced, and irreducible tensors make up a free R-module $\bigotimes^\bullet_{\text{irr}} \mathfrak{g}$. The universal enveloping algebra $U(\mathfrak{g})$ of $\mathfrak{g}$ is by definition the quotient of $\bigotimes^\bullet \mathfrak{g}$ by the two-sided ideal I generated by $x \otimes y - y \otimes x - [x,y]$ for x, y in $\mathfrak{g}$.

3.1. Theorem. (Poincaré-Birkhoff-Witt) *The images of the reduced monomials form a basis of $U(\mathfrak{g})$.*

In other words,

$$\bigotimes^\bullet \mathfrak{g} = \bigotimes^\bullet_{\text{irr}} \mathfrak{g} \oplus I .$$

I should remark here that choosing $\geq$ rather than $\leq$ is somewhat arbitrary. There is, as far as I can tell, no universally adopted convention. My choice has been made with later developments in mind.

Proof. It has to be proven that (a) every element of each $\bigotimes^\bullet \mathfrak{g}$ is equivalent modulo I to a linear combination of ordered monomials and (b) this linear combination is unique.

(a) Because the expression $x \otimes y - y \otimes x - [x,y]$ is bilinear and because it changes sign if x and y are swapped:

3.2. Lemma. *The ideal I is the two-sided ideal generated by $x \otimes y - y \otimes x - [x, y]$ for $x < y$ in X.*

For monomials A, B and $x < y$ in X, define the operator $\sigma = \sigma_{A,x\otimes y,B}$ on $\bigotimes^{\bullet}\mathfrak{g}$ to be that which takes

$$A \otimes x \otimes y \otimes B \longmapsto A \otimes (x \otimes y - [x, y]) \otimes B$$

and fixes all other monomials. Thus $\sigma(\alpha) - \alpha$ lies in I. If $x < y$ the tensor $A \otimes (y \otimes x + [x, y]) \otimes B$ is called a reduction of $A \otimes x \otimes y \otimes B$, and such operators are also called reductions. A tensor is irreducible if and only if it is fixed by all reductions.

Claim (a) means that by applying a finite number of reductions any tensor can be brought to an irreducible one. The basic idea is straightforward. Start with a tensor α. As long as α possesses monomials that are not reduced, replace one of them by the appropriate reduction. Every reduction in some sense simplifies a monomial, so it is intuitively clear that the process must stop. The problem in this description is that reduction triggers a cascade of others that is difficult to predict. Also, what exactly is simplification?

If $A = x_1 \otimes \ldots \otimes x_n$, let $|A| = n$ and let $\ell(A)$ be the number of 'inversions' $x_i < x_j$ with $i < j$. The monomials A with $\ell(X) = 0$ are the reduced ones, so $\ell(A)$ is a measure of how far A is from being reduced. If A and B are two monomials, then we say $A \prec B$ if either $|A| < |B|$ or $|A| = |B|$ but $\ell(A) < \ell(B)$. If $|A| = k$ then $\ell(A) \le k(k - 1)/2$, and from this it follows that if $|A| = n$ the length of any chain

$$A = A_1 \succ A_2 \succ \ldots$$

is at most $n(n - 1)/2 + (n - 1)(n - 2)/2 + \cdots = n(n^2 - 1)/6$.

If α is a tensor, defines its **support** $\operatorname{supp}(\alpha)$ to be the set of monomials appearing with non-zero coefficients in its expression. To each tensor $a = \sum c_X X$ we can associate a graph. Its nodes will be defined recursively: (a) the monomials in the support of α are nodes; (b) if X is in it, then the support of each reduction of X are nodes. From each node $X = A \otimes x \otimes y \otimes B$ with $x < y$ there is a directed edge from X to each node in the reduction $\sigma_{A,x\otimes y,B}(X)$, labeled by $(A, x \otimes y, B)$. If any sequence of reduction operators is applied to α, the support of the result is contained in the set of nodes of this graph (but may well be a proper subset, because of possible cancellation).

Let $\Gamma = \Gamma_\alpha$ be this graph, and let S be the set of all inverted pairs $(A, x \otimes y.B)$ with $x < y$. From each node we have a set of directed edges indexed by a certain subset of S. Define for each s in S a map r_s from subsets of the nodes to other subsets of nodes, taking Θ to the union of the targets of edges labeled by s in Θ. Every directed path in Γ has finite length. In particular, ther are no loops in Γ.

3.3. Lemma. *In this situation, every infinite composition of maps r_s is stationary.*

Proof. Because if not there would exist an infinite directed path.

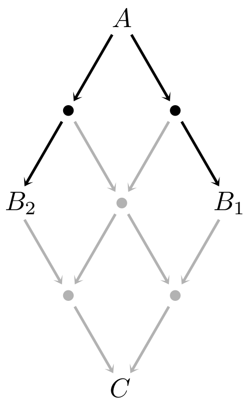
Therefore any sequence of reductions will eventually be stationary. If we follow the principle of applying a reduction as long as some remaining monomial is not treduced, the process will terminate with an irreducible tensor. Thus we have proved claim (a). That's the easy half of the Theorem.

(b) It remains to show uniqueness. The previous argument showed that there always exists a series of reductions leading to an irreducible tensor. Call an element of $\bigotimes^{\bullet}\mathfrak{g}$ **uniquely reducible** if every chain of reductions that ends in an irreducible expression ends in the same expression. In order for the irreducible tensors to be linearly independent, it is necessary and sufficient that all elements be uniquely reducible.

The following is a step towards proving this.

3.4. Lemma. (PBW Confluence) *If A is a monomial with two simple reductions $A \to B_1$ and $A \to B_2$ then there exist further reductions $B_1 \to C$ and $B_2 \to C$.*

The term 'confluence' seems to have its origin in [Newman:1942], where a similar question is discussed for purely combinatorial (as opposed to algebraic) objects. In the cases discussed by Newman, this is called the Diamond Lemma, and the following picture suggests straightforward consequences for composites of reductions.



The complication in the case at hand is that successive reductions can cancel out terms from early ones. That is exactly the problem that [Bergman:1978] is concerned with.

Proof. Suppose that A is a monomial with two simple reductions $\sigma\colon A \to B_1$ and $\tau\colon A \to B_2$. We must find further reductions $B_1 \to C$ and $B_2 \to C$.

If the reductions are applied to non-overlapping pairs there is no problem. An overlap occurs for a term $A \otimes x \otimes y \otimes z \otimes B$ with $x < y < z$. It gives rise to a choice of reductions.

$$\begin{aligned} x \otimes y \otimes z &\longrightarrow y \otimes x \otimes z + [x, y] \otimes z \\ x \otimes y \otimes z &\longrightarrow x \otimes z \otimes y + x \otimes [y, z] \,. \end{aligned}$$

But then

$$\begin{aligned} y \otimes x \otimes z + [x, y] \otimes z &\longrightarrow y \otimes z \otimes x + y \otimes [x, z] + [x, y] \otimes z \\ &\longrightarrow z \otimes y \otimes x + [y, z] \otimes x + y \otimes [x, z] + [x, y] \otimes z \\ x \otimes z \otimes y + x \otimes [y, z] &\longrightarrow z \otimes x \otimes y + [x, z] \otimes y + x \otimes [y, z] \\ &\longrightarrow z \otimes y \otimes x + z \otimes [x, y] + [x, z] \otimes y + x \otimes [y, z] \end{aligned}$$

and the difference

$$\begin{aligned} &([y, z] \otimes x - x \otimes [y, z]) + (y \otimes [x, z] - [x, z] \otimes y) + ([x, y] \otimes z - z \otimes [x, y]) \\ &\qquad = ([y, z] \otimes x - x \otimes [y, z]) + ([z, x] \otimes y - y \otimes [z, x]) + ([x, y] \otimes z - z \otimes [x, y]) \end{aligned}$$

between the right hand sides lies in I because of Jacobi's identity. ∎

We'll see confluence again later on.

If u is any uniquely reducible element of $\bigotimes^{\bullet} \mathfrak{g}$, let $\operatorname{irr}(u)$ be the unique irreducible element it reduces to.

3.5. Lemma. *If u and y are uniquely reducible, so is $u + v$, and $\operatorname{irr}(u + v) = \operatorname{irr}(u) + \operatorname{irr}(v)$.*

Proof. Let τ be a composite of reductions such that $\tau(\alpha + \beta) = \gamma$ is irreducible. Since α is uniquely reducible, there exists some σ such that $\sigma(\tau(\alpha) = \operatorname{irr}(\alpha)$. Since β is irreducible, there exists ρ such that $\rho(\sigma(\tau(\beta))) = \operatorname{irr}(\beta)$. Since all reductions fix irreducible tensors and reductions are linear

$$\begin{aligned} \rho(\sigma(\tau(\alpha + \beta))) &= \gamma \\ &= \rho(\sigma(\tau(\alpha) + \rho(\sigma(\tau(\beta))) \\ &= \operatorname{irr}(\alpha) + \operatorname{irr}(\beta) \,. \end{aligned}$$

∎

Now for the proof of the PBW Theorem. We want to show that every tensor is uniquely-irreducible. By this Lemma it suffices to show that every monomial X is uniquely irreducible. We do this by induction on $|X|$. For $|X| \le 2$ the claim is immediate.

Suppose $|X| \ge 3$. Suppose we are given two chains of reductions, taking X to Z_1 and Z_2. let $X \to Y_1$ and $X \to Y_2$ be the first steps in each. By Lemma 3.4 there exists further composite reductions $Y_i \to W$. By the induction hypothesis on the W_i we know that $\mathrm{irr}(W) = \mathrm{irr}(Y_i) = Z_i$ for each i. □

The notion of confluence used here is a tool of great power in finding normal forms for algebraic expressions. It is part of the theory of **term rewriting** and **critical pairs**, and although it has been used informally for a very long time, the complete theory seems to have originated in [Knuth-Bendix:1965]. It has been rediscovered independently a number of times. A fairly complete bibliography as well as some discussion of the history can be found in [Bergman:1977].

4. Free Lie algebras and tensor products

I now prove the propositions stated in the introduction. I follow the exposition in [Serre:1965] very closely.

4.1. Theorem. *The canonical map from $U(\mathcal{L}_X)$ to $\bigotimes^\bullet R_X$ is an isomorphism.*

Proof. The map from R_X to $\mathcal{L}_X$ is an embedding, and by Poincaré-Birkhoff-Witt this in turn embeds into $U(\mathcal{L}_X)$ and therefore induces a ring homomorphism from $\bigotimes^\bullet R_X$ to $U(\mathcal{L}_X)$. This is inverse to that from $U(\mathcal{L}_X)$ to $\bigotimes^\bullet R_X$. □

It remains now to prove:

4.2. Theorem. *If $|X| = d$, then*

$$\ell_X(n) = \left(\frac{1}{n}\right) \sum_{m|n} \mu(m) d^{(n/m)} .$$

4.3. Theorem. *The free Lie algebra $\mathcal{L}_X$ is a free R-module, as are all its homogeneous components $\mathcal{L}_X^{(n)}$.*

4.4. Theorem. *The canonical map from $\mathcal{L}_X$ to L_X is an isomorphism.*

Proof of all of these. I first point out that Theorem 4.4 follows from Theorem 4.3. The map from $\mathcal{L}_X$ to L_X is certainly surjective, and it is injective by Theorem 4.1, since the Poincaré-Birkhoff-Witt theorem implies that $\mathcal{L}_X$ embeds into $U(\mathcal{L}_X)$. In particular, Theorem 4.4 is true if R is a field.

But if $\mathcal{L}_X$ is R-free, the formula for $\ell_d(n)$ is also easy to see. If $d = |X|$, then the dimension of $T_n = \bigotimes^n R_X$ is d^n, and we have the formula for the generating function

$$\sum_0^\infty (\mathrm{rank}\ T_n)\, t^n = \sum_0^\infty d^n t^n = \frac{1}{1-dt} .$$

The identification of $U(\mathcal{L}_X)$ with $\bigotimes^\bullet R_X$ gives us another formula for the same function. Choose a basis λ_i of $\mathcal{L}_X$ and a linear ordering of it (or, effectively, its set of indices I). Let $d_i = |\lambda_i|$. The Poincaré-Birkhoff-Witt theorem implies that a basis of $U(\mathcal{L}_X)$ is made up of the ordered products

$$\lambda^e = \prod_{i \in I} \lambda_i^{e_i}$$

with $e_i = 0$ for all but a finite number of i. The degree of this product is $\sum e_i d_i$. We then also have

$$\sum_0^\infty (\mathrm{rank}\ T_n)\, t^n = \prod_1^\infty \frac{1}{(1-t^m)^{\ell_d(m)}} .$$

Therefore

$$\prod_1^\infty \frac{1}{(1-t^m)^{\ell_d(m)}} = \frac{1}{1-dt}$$

which implies that

$$d^n = \sum_{m|n} m\ell_d(m) \,,$$

and by Möbius inversion (§16.4 of [Hardy-Wright:1960]) the formula in Theorem 4.2.

Now suppose $R = \mathbb{Z}$. The case just dealt with shows that the dimension $\mathbb{Z}/p \otimes \mathcal{L}_X^{(n)}$ is independent of p, which in turn implies that $\mathcal{L}_X^{(n)}$ is free over $\mathbb{Z}$.

The case of arbitrary R is immediate, since $\mathcal{L}_{X,R} = R \otimes \mathcal{L}_{X,\mathbb{Z}}$. This concludes all pending proofs.

Here are the first few dimensions for $d = 2$:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
d_n	2	1	2	3	6	9	18	30	56	99	186	335	630	1161	2182	4080

The proofs of these theorems are highly non-constructive. These problems now naturally arise:

- *specify an explicit basis for all of $\mathcal{L}_X$;*
- *find an algorithm to do explicit computations in terms of it;*
- *make a more explicit identification of $U(\mathcal{L}_X)$ with $\bigotimes^\bullet R_X$.*

These problems will all be solved in terms of *Hall sets*.

5. Hall sets—motivation

Continue to let X be any finite set, $V = R_X$ the free R-module with basis X. Inside $\bigotimes^\bullet V$ lies the Lie algebra L_X generated by V and the bracket operation $[P, Q] = PQ - QP$. This is spanned by the **Lie polynomials** generated by elements of X. Recursively, we have these generating rules: (a) if x is in X then its image in $\bigotimes^\bullet V$ is a Lie polynomial and (b) if P and Q are Lie polynomials so is $[P, Q]$.

If $X = \{x, y\}$, some of the first few Lie polynomials, in both bracket and tensor form, are

$$\begin{aligned}
&x \\
&y \\
[x, y] &= x \otimes y - y \otimes x \\
[x, [x, y]] &= x \otimes (x \otimes y - y \otimes x) - (x \otimes y - y \otimes x) \otimes x \\
&= x \otimes x \otimes y - 2x \otimes y \otimes x + y \otimes x \otimes x \\
[y, [x, y]] &= y \otimes (x \otimes y - y \otimes x) - (x \otimes y - y \otimes x) \otimes y \\
&= 2y \otimes x \otimes y - y \otimes y \otimes x - x \otimes y \otimes y \,.
\end{aligned}$$

A bracket expression involving elements of X, when expanded into tensors, always becomes a sum of tensor monomials all of the same degree as the original expression.

The embedding of X into V determines a map from $\mathcal{L}_X$ to L_X, which we know to be an isomorphism. It determines also a ring isomorphism of $U(\mathcal{L}_X)$ with $\bigotimes^\bullet V$. Any monomial gives rise to a Lie polynomial in the tensor algebra as well as an element of $\mathcal{L}_X$.

The isomorphism of $U(\mathcal{L}_X)$ with $\bigotimes^\bullet V$ is more than a little obscure. It is not at all easy to identify elements of L_X when expressed in tensor form. This is a problem that we shall see in many guises. Here, I shall sketch quickly one way to implement the Poincaré-Birkhoff-Witt Theorem for $U(\mathcal{L}_X)$ inside $\bigotimes^\bullet V$. That is to say, I'll show how to express any tensor as a linear combination of ordered monomials in certain elements of $\mathcal{L}_X$. Temporarily, in order to simplify terminology, I'll call such an expression a PBW polynomial.

The matter of ordering is easy to deal with. First of all order X. If $X = \{x, y\}$, for example, I could set $x < y$. Next, to each Lie polynomial is associated its word. For example, the word of $[[x, [y, x]], x]$ is $xyxx$. I'll put a weak order on Lie monomials be saying $p \le q$ is the word of p comes before q in a dictionary. For example, $[x, y]$ comes before $[[x, [y, x]], x]$, which comes in turn before $[[y, [y, x]], x]$. This is not a strict ordering, since several Lie polynomials might have the same word, but I'll not pat attention to that problem.

The algorithm I am about to explain will find an expression for any element of $\bigotimes^\bullet V$ as a PBW polynomial. The ultimate output might not be so clear, but the algorithm nonetheless is quite precise. As for the output, that should become clearer in time, but it should be clear that it has something to do with PBW.

This algorithm will be linear, so in order to describe, I may it start with a tensor monomial, say $x_1 \otimes \ldots x_n$. At any point in the process, we will be facing a linear combination of products $p_1 \otimes \ldots \otimes p_n$ of Lie polynomials. We scan this term by term from the left to find the first inversion. If $p_1 \ge \ldots \ge p_n$, we leave this as it is. Otherwise we shall have a product in which either (1) $p_1 \le p_2$ or (2) $p_{k-1} \ge p_k < p_{k+1}$ for some $k \ge 2$. We replace these terms in the product by a sum

$$p_2 \otimes p_1 + [p_1, p_2] \ \text{ or } \ p_{k-1} \otimes p_{k+1} \otimes p_k + p_{k-1} \otimes [p_k, p_{k+1}] \,.$$

We keep doing this until all terms are products of weakly decreasing sequences of Lie polynomials. There might well be some ambiguity in what the point is, but at least the process itself is unambiguous.

For example, here is a sample sequence:

$$\begin{aligned}
&\boldsymbol{x} \otimes \boldsymbol{x} \otimes \boldsymbol{y} \\
&\boldsymbol{x} \otimes \boldsymbol{y} \otimes \boldsymbol{x} + x \otimes [x, y] \\
&y \otimes x \otimes x + [x, y] \otimes x + \boldsymbol{x} \otimes [\boldsymbol{x}, \boldsymbol{y}] \\
&y \otimes x \otimes x + [x, y] \otimes x + [\boldsymbol{x}, \boldsymbol{y}] \otimes \boldsymbol{x} + [x, [x, y]] \\
&y \otimes x \otimes x + 2\left([x, y] \otimes x\right) + [x, [x, y]] \,.
\end{aligned}$$

There are a several questions that arise immediately. First of all, *inside a term there can be several inversions pairs requiring to be swapped. I have specified which is to be dealt with first, but does it matter which we choose?* Second, *can we characterize the Lie polynomials we get in the end?* Third, *to what extent is this really implementing PBW?* This is a legitimate question,because I have not specified, as one does in PBW, a linear order on Lie polynomials.

We can get some idea of an answer to the second of these by examining again what happens when Lie polynomials are created in the process. I recall that if we are facing

$$p_k \otimes p_{k+1} \quad (p_k < p_{k+1}) \,,$$

which changes to

$$p_{k+1} \otimes p_k + p_{k-1} \otimes [p_k, p_{k+1}] \,.$$

So the first condition is that if $[p, q]$ is a Lie polynomial created by this process, then $p < q$. A second that is a bit more difficult to explain. How can a Lie polynomial $[p, [q, r]]$ arise? Let's look at one example where it does. Suppose we encounter a product $p \otimes q \otimes r$ with $p \ge q \le r$. This has become $p \otimes (r \otimes q) + p \otimes [q, r]$. But then to get the bracket $[p, [q, r]]$ we must have $p < [q, r]$. In other words, p must be in a sandwich:

$$q \le p < [q, r] \,.$$

It turns out that we now have in hand exactly the necessary and sufficient conditions for a Lie polynomial $[p, q]$ to appear in the process we have described above: (a) $p < q$; (b) if $q = [r, s]$ is compound, then $r \le p$.

As we shall see, these conditiosn characterize a very important set of Lie polynomis. Among other things, it will turn out that such polynomials are distinguished by their words, so the ambiguity in order does not in fact matter. This process, in other words, has a certain magical quality—we seem to get more out of it than we can reasonably expect.

The conditions (a) and (b) are those which, when applied recursively, characterize the basis elements of $\mathcal{L}_X$ called **Hall monomials** (with respect to dictionary order). And the process described above then does implement PBW with respect to that basis and that order. In defining Hall monomials precisely in the next section, we shall place ourselves in a somewhat more general situation. A number of pleasant things will happen, though—not only shall we find explicit bases for $\mathcal{L}_X$, but we shall also find several practical algorithms for dealing with them. One useful fact is that in the final reduction process, which is not very different from the one described here, it will be admissible to invert any pair in decreasing order, not necessarily just the first one occurring. The final expression will therefore be independent of exactly what order inversions are done. This will be a second example of confluence, which we have seen before in the proof of PBW.

6. Hall sets—definitions

A **Hall set** in M_X is a set H of monomials in a magma together with a linear order on H satisfying certain conditions. The first is on the order alone:

(1) for $\lfloor p, q \rfloor$ in H, $p < \lfloor p, q \rfloor$.

An order satisfying this condition is called a **Hall order**. The other conditions are recursive:

(2) the set X is contained in H;
(3) the monomial $\lfloor p, q \rfloor$ is in H if and only if both (a) p and q are both in H with $p < q$; and (b) either (i) q lies in X or (ii) $q = \lfloor r, s \rfloor$ and $r \le p$.

I. e. p is sandwiched in between r and q. With a given X and even a given order on X, there are many, many different Hall sets, because there are many possibilities for Hall orders. The elements of a Hall set are often called **Hall monomials**, but occasionally **Hall trees** to emphasize their structure as graphs. A Hall tree has the recursive property that the descendants of any node of a Hall tree make up a Hall tree. By induction, the conditions on Hall trees can be put more succinctly as the conditions

$$h_{*0} < h_{*1}, \quad h_{*0} < h_*$$

for any compound node h_* of h, and

$$h_{*10} \le h_{*0}$$

when h_{*1} is also compound.

In reading the literature, one should be aware that there are many variant definitions of Hall sets, mainly corresponding to a reversal of orders. Instead of the condition (3) that $p < \lfloor p, q \rfloor$, the original definition of [Hall:1950] imposed on the order the stronger condition that if $|p| < |q|$ then $p < q$. This is also the one found in [Serre:1965]. The definition here is less restrictive than that of Hall, and is usefully more flexible. For long time after Hall's original paper there was some confusion as to exactly what properties of a Hall set were necessary. Although different orders produce distinct Hall sets, all of them give rise to good bases of $\mathcal{L}_X$. The situation was made much clearer when J. Michel and X. G. Viennot independently realized that the condition $p < \lfloor p, q \rfloor$ was sufficient for all practical purposes, and [Viennot:1978] showed that in the presence of (2) it was also necessary.

Much of my exposition is taken from [Reutenauer:1993], but my conventions a bit are different from his. There are in fact several different valid conventions. At the end of Chapter 4 of Reutenauer's book is an illuminating short discussion of all of the ones in use, as well as a short account of the somewhat intricate history. Incidentally, the paper first explicitly defining a Hall set is the one by Marshall Hall, but the idea originated in [Hall:1934], which is by the English group theorist Philip Hall. His article is about commutators in p-groups and at first glance has nothing to do with Lie algebras. It was [Magnus:1937] who brought them in. The book [Serre:1965] explains nicely the close connection between free Lie algebras and free groups.

The definition suggests how to construct Hall sets, at least as many Hall monomials as one wants. One starts with the set X as well as an order on it, then proceeds from one degree to the next. To find all Hall monomials of order n given those of smaller order one adds first all $\lfloor p, x \rfloor$ with $p < x$ of order $n-1$; then for each compound $q = \lfloor r, s \rfloor$ of order $k < n$ one adds all the $\lfloor p, q \rfloor$ with p of order $n-k$ such that $r \le p < q$. Finally, one inserts the newly generated monomials of degree n into an ordering, maintaining the condition that $p < \lfloor p, q \rfloor$. As we'll see later, a rough estimate of $|H_n|$ is d^n/n if $|X| = d$, a number that grows rapidly.

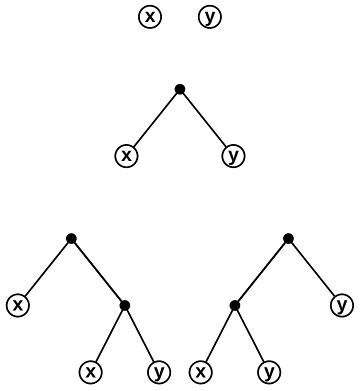
The order is a non-trivial part of the definition. All that's needed is a strict linear order on H itself. This is often an order induced from an order, not necessarily strict, on all of M_X. The order on M_X is usually taken also to be a Hall order. But it is possible to construct a Hall set and choose an order dynamically, as Serre's book does. In this process, assign an arbitrary order to the new monomials of degree n as they are found, and then specify that $p < q$ if $|p| < |q| = n$. This is conceptually simple if not very satisfying. Other ways to proceed depend on results to be proven later on. The words composed of elements of X can be ordered **lexicographically**, that is to say by dictionary order. If p and q are words then $p < q$ if either (a) p is an initial word of q or (b) the first letter of p different from the corresponding letter of q is less than it. Thus if $x < y$

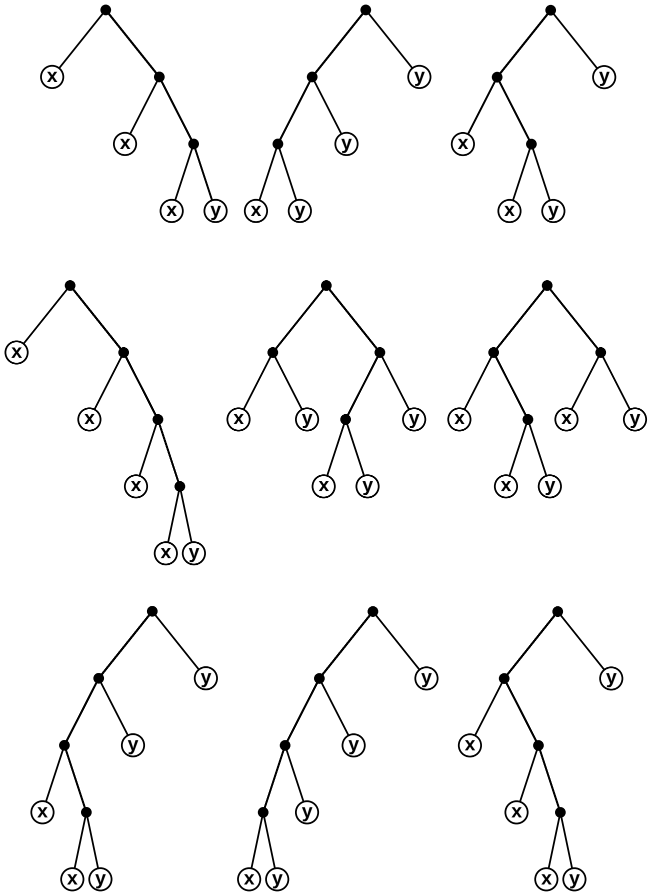
$$x < xx < xxxy < xxy < y < yx\,.$$

We'll see later that Hall monomials are uniquely determined by their words. This means that one possible linear order on a Hall set is the lexicographic order on their words. With this choice, here is a list of all Hall words for the set $X = \{x, y\}$ up to order 5 (with $x < y$):

$$\begin{array}{ll}
H_1\colon & x,\ y \\
H_2\colon & \lfloor x, y \rfloor \\
H_3\colon & \lfloor x, \lfloor x, y \rfloor \rfloor,\ \lfloor \lfloor x, y \rfloor, y \rfloor \\
H_4\colon & \lfloor x, \lfloor x, \lfloor x, y \rfloor \rfloor \rfloor,\ \lfloor \lfloor x, \lfloor x, y \rfloor \rfloor, y \rfloor,\ \lfloor \lfloor \lfloor x, y \rfloor, y \rfloor, y \rfloor, \\
H_5\colon & \lfloor x, \lfloor x, \lfloor x, \lfloor x, y \rfloor \rfloor \rfloor \rfloor,\ \lfloor \lfloor x, \lfloor x, \lfloor x, y \rfloor \rfloor \rfloor, y \rfloor,\ \lfloor \lfloor x, \lfloor x, y \rfloor \rfloor, \lfloor x, y \rfloor \rfloor \\
 & \qquad \lfloor \lfloor \lfloor x, \lfloor x, y \rfloor \rfloor, y \rfloor, y \rfloor,\ \lfloor \lfloor x, y \rfloor, \lfloor \lfloor x, y \rfloor, y \rfloor \rfloor,\ \lfloor \lfloor \lfloor \lfloor x, y \rfloor, y \rfloor, y \rfloor, y \rfloor
\end{array}$$

These can be pictured:





The principal result about Hall sets—and indeed their *raison d'etre*—is that the image of a Hall set in $\mathcal{L}_X$ is a basis. We'll see eventually a constructive proof, as well as an algorithm for calculating Lie brackets in terms of this basis.

7. Admissible sequences

The words associated to Hall monomials are called **Hall words**. Of course going from a tree to its word loses a lot of structure, and it is therefore somewhat satisfying that, as we shall see shortly, the map from Hall monomials to their words is injective. *Is there a good way to tell if a given word is a Hall word? How can one reconstruct a Hall tree from its word?* Answering these will involve a closer examination of the algorithm I exhibited earlier for making PBW explicit.

That answer is in terms of **admissible sequence** and transformations of them called **packing** and **unpacking**. An admissible sequence of Hall monomials is a sequence

$$h_1, h_2, \dots, h_n \quad (h_i \in H)$$

with the property that for each k we have either (a) h_k lies in X or (b) h_k is compound and $h_{k,0} \leq h_i$ for all $i < k$.

The following is elementary:

7.1. Lemma. *Any sequence of elements of X is an admissible sequence, and any weakly decreasing sequence of Hall monomials is admissible.*

Any segment of an admissible sequence is also admissible. Adding a tail of elements in X to an admissible sequence always produces an admissible sequence.

7.2. Lemma. *If we are given an admissible sequence with $h_i \ge h_k$ for $i < k$ but $h_k < h_{k+1}$ then the sequence we get from it by replacing the pair h_k, h_{k+1} by the single monomial $\lfloor h_k, h_{k+1} \rfloor$ is also admissible.*

The proof is straightforward.

In general, if h_i is an admissible sequence and replacing a pair h_k, h_{k+1} by the new monomial $\lfloor h_k, h_{k+1} \rfloor$ is still an admissible sequence of Hall monomials, I say the new one is obtained from the old one by **packing**. If σ is the original sequence and τ the new one, I write $\sigma \rightarrow \tau$. If τ is obtained from σ by zero or more such replacements, I write $\sigma \overset{*}{\rightarrow} \tau$. According to the Lemma, the only admissible sequences that cannot be packed are the weakly decreasing ones. As an example, here is how to pack $xxyxy$:

$$\begin{array}{l} x \ge x < y \quad x \quad y \\ x < \lfloor x, y \rfloor \quad x \quad y \\ \lfloor x, \lfloor x, y \rfloor \rfloor \ge x < y \\ \lfloor x, \lfloor x, y \rfloor \rfloor < \lfloor x, y \rfloor \\ \lfloor \lfloor x, \lfloor x, y \rfloor \rfloor, \lfloor x, y \rfloor \rfloor \end{array}$$

packs the sequence $xxyxy$ into a single Hall monomial. Packing does not require an initial weakly decreasing segment, however, and there might be more than one way to pack an admissible sequence.

Unpacking is the reverse process, which replaces a monomial $\lfloor h_a, h_b \rfloor$ by the pair h_a, h_b if the result is admissible. The following Lemma asserts that this is always possible unless the sequence consists only of elements of X.

7.3. Lemma. *Suppose τ to be an admissible sequence of Hall monomials. If h_k is the compound monomial in the sequence furthest to its right, then the sequence σ we get by unpacking h_k to $h_{k,0}$, $h_{k,1}$ is again admissible.*

Thus $\sigma \rightarrow \tau$. As a consequence, any admissible sequence of Hall monomials, and in particular any weakly decreasing sequence, can be completely unpacked to a sequence of elements of X. If we pack a word and then unpack it, we get the original word back again because it is the word corresponding to the concatenation of the trees. What is not so obvious is that every complete chain of packing operations starting from the same word always gives the same final result. In other words, what we know at the moment is that any sequence of letters in X packs to a weakly decreasing sequence of Hall monomials, and conversely any weakly decreasing sequence of Hall monomials unpacks to a sequence of letters. What we want to know now is that the sequence of words determines the sequence of Hall monomials. This will follow from:

7.4. Lemma. (Packing confluence) *If $\rho \overset{*}{\rightarrow} \sigma_1$ and $\rho \overset{*}{\rightarrow} \sigma_2$ then there exists τ with $\sigma_1 \overset{*}{\rightarrow} \tau$ and $\sigma_2 \overset{*}{\rightarrow} \tau$.*

Proof. As earlier, where I referred to the diamond diagram, this reduces to the simplest case where $\rho \rightarrow \sigma_1$ and $\rho \rightarrow \sigma_2$. Suppose the first packs h_i, h_{i+1} and the second h_j, h_{j+1}. We may assume $i < j$. Then necessarily $i + 1 < j$ and we may pack both pairs simultaneously into an admissible sequence τ which both σ_1 and σ_2 pack into. ▮

Consequently:

7.5. Proposition. *Every word in X can be expressed uniquely as a concatenation of Hall words associated to a weakly decreasing sequence of Hall monomials.*

This is very closely related to the PBW reduction process applied to $U(\mathcal{L}_X)$. But it also leads to the following important result, which is a special case where we start with the word of a Hall tree:

7.6. Corollary. *Packing is a bijection between words of length n and weakly decreasing sequences of Hall monomials of total degree n.*

It might be useful to have a list of the assignments for length 3:

$$\begin{aligned} xxx&\colon\ xxx \\ xxy&\colon\ [x,[x,y]] \\ xyx&\colon\ [x,y]x \\ xyy&\colon\ [[x,y],y] \\ yxx&\colon\ yxx \\ yxy&\colon\ y[x,y] \\ yyx&\colon\ yyx \\ yyy&\colon\ yyy \end{aligned}$$

7.7. Corollary. *The map from H to the set of words in X that takes h to $\omega(h)$ is injective.*

8. Hall words for lexicographic order

One consequence of Corollary 7.7 is the claim made earlier, that lexicographic order defines a strict order on Hall monomials. It certainly satisfies the condition that $p < \lfloor p, q \rfloor$, and now we know that if $p \neq q$ then either $p < q$ or $q < p$. A **lexicographic Hall word** is a word associated to a Hall monomial in a Hall set defined with respect to lexicographic order. These are also called **Lyndon words**. Given any Hall set, the packing algorithm described earlier will tell us whether a word is a Hall word associated to it. Lexicographic Hall words possess a simple if surprising characterization in terms of the words alone. It is useful in manual computations, and also of theoretical importance.

Let $\mathcal{W}$ be the set of all words w which are lexicographically less than their proper tails. Thus w is in $\mathcal{W}$ if and only if $w < v$ whenever $w = uv$ is a factorization with $u \neq \emptyset$. In particular $X \subset \mathcal{W}$.

8.1. Theorem. *A string w is a lexicographic Hall word if and only if it lies in $\mathcal{W}$.*

It is a worthwhile exercise to check that the words associated to the Hall monomials of length at most 5, which I have exhibited earlier, are all in $\mathcal{W}$. For example, $xxyxy$ is in $\mathcal{W}$ since

$$\begin{aligned} xxyxy &< xyxy \\ &< yxy \\ &< xy \\ &< y\,. \end{aligned}$$

Proof. We need two lemmas. The first is straightforward.

8.2. Lemma. *If $u < v$ are two words with u less than v in lexicographic order, then $wu < wv$ for all words w. If u is not a prefix of v, then so is $uw < vw$.*

8.3. Lemma. *If $x < y$ are in $\mathcal{W}$, so is xy. Conversely, suppose $w = uv$ lies in $\mathcal{W}$, with v least among its proper tails. Then both u and v are in $\mathcal{W}$ and $u < v$.*

Proof. Let $z = xy$ with both x, y in $\mathcal{W}$, and suppose $z = uv$ to be a proper factorization. It is to be shown that $z < v$.

Suppose $\ell(u) < \ell(x)$. Then there exists a proper factorization $x = uu_*$. Since $z = uu_*y$, we must have $v = u_*y$. Since x is in $\mathcal{W}$, $x < u_*$. Since x is not a prefix of u_*, we conclude that $z = xy < u_*y = v$.

Suppose $\ell(u) = \ell(x)$. Then in fact $u = x$ and $v = y$. By assumption, $x < y$. If x is not a prefix of y then $z = xy < y$ also, since x and y, hence also z and y, differ in the first character where they are not equal. If x is a prefix of y, say $y = xy_*$. Since y is in $\mathcal{W}$, $y < y_*$ and then also $xy < xy_*$.

Finally, suppose $\ell(u) > \ell(x)$. Then $u = xu_*$, $y = u_*v$ are proper factorizations. But then $y < v$, and by the previous case $z < y$, so $z < v$. This concludes the proof of the first half of the Lemma.

Now suppose that w lies in $\mathcal{W}$. Let $w = uv$ with v the least among the proper right factors of w. I must show that u and v are both in $\mathcal{W}$ and that $u < v$. That v lies in $\mathcal{W}$ is an immediate consequence of its specification.

We now want to show that u is in $\mathcal{W}$. Suppose $u = u_*v_*$, and suppose $u > v_*$. (a) If v_* is not a prefix of u, it follows that $w = uv > v_*v$, contradicting that w lies in $\mathcal{W}$. (b) If v_* is a prefix of u, say $u = v_*u_{**}$, then since u is in $\mathcal{W}$ we have $u = v_*u_{**}v < v_*v$. This implies that $u_{**}v < v$, contradicting the choice of v. So $u < v_*$ and u lies in $\mathcal{W}$.

Since $uv < v$ and uv is not a prefix of v, there exists $k \le |v|$ such that $(uv)_i = v_i$ for $i < k$ but $(uv)_k < v_k$. If $k > |u|$ then u is a prefix of v; otherwise $(uv)_k = u_k < v_k$. In either case $u < v$. (This argument shows that $u < v$ for any factorization $w = uv$ of w in $\mathcal{W}$.) ∎

Now for the proof of Theorem 8.1. It is now an easy induction to see that $\omega(p)$ lies in $\mathcal{W}$ for all Hall monomials. If $|p| = 1$ this is trivial. If $p = \lfloor q, r \rfloor$ then $\omega(p) = \omega(q)\omega(r)$ and $\omega(p)$ is in $\mathcal{W}$ by induction and Lemma 8.3.

So now suppose that w lies in $\mathcal{W}$. I am going to explain how to find explicitly a Hall monomial whose word is w.

Step 1. If $|w| = 1$, return w itself, an element of X.

Step 2. Otherwise, according to Lemma 8.3 there exists at least one proper factorization $w = uv$ with $u < v$ and both u and v in $\mathcal{W}$. Choose such a factorization with $|v|$ minimal.

Step 3. By recursion, $u = \omega(p)$ and $v = \omega(q)$ for Hall monomials p, q. If $|q| = 1$ then $\lfloor p, q \rfloor$ is a Hall monomial, and its word is w. Otherwise $q = \lfloor q_0, q_1 \rfloor$ with $q_0 < q_1$ both Hall monomials. If $q_0 \le p$ then $\lfloor p, q \rfloor$ is a Hall monomial.

Step 4. Otherwise $p < q_0$. In this case $\omega(p)$ and $\omega(q_0)$ lie in $\mathcal{W}$, so by Lemma 8.3 so is the product $\omega(p)\omega(q_0)$. Since q_1 is a Hall monomial, so is $\omega(q_1)$ in $\mathcal{W}$. Since w_* is a prefix of w, we know that $w_* < w$. By assumption $w < \omega(q_1)$, so $p_* < q_1$. Since $|q_1| < |q| = |v|$, this contradicts the choice of v as of minimal length. ∎

For w in $\mathcal{W}$, a factorization $w = uv$ with u, v also in $\mathcal{W}$ is called a **Lyndon factorization**. Such factorizations are not necessarily unique. For example, if $\lfloor p, q \rfloor$ is a lexicographic Hall monomial for $X = \{x, y\}$, then $\lfloor \lfloor p, q \rfloor, z \rfloor$ is one for $\{x, y, z\}$. But the Lyndon word $\omega(p)\omega(q)z$ has Lyndon factorizations

$$(\omega(p)\omega(q))\, z \text{ and } \omega(p)(\omega(q)\, z)\,.$$

It is the first that gives originates from a Hall monomial, in agreement with the argument just presented sinc$z$ is shorter than $\omega(q)\, z$.

8.4. Proposition. *If $u\,v$ and $v\,w$ are non-empty lexicographic Hall words then so is $u\,v\,w$.*

Proof. Since uv and vw are Hall words,

$$u < uv < v < vw < w\,.$$

From here, it is straightforward to see that uvw lies in $\mathcal{W}$. ∎

If we start with a Hall monomial f in $\bigotimes^\bullet R_X$ and write out its full tensor expansion, we get a linera combination of tensor monomials. All are of the same degree, equal to $|f|$. There is an obvious bijection between tensor products and words.

8.5. Proposition. *In the tensor expansion of a Hall monomial f, the least term is that associated to $\omega(f)$.*

Proof. Follows from the reduction algorithm in $\bigotimes^\bullet R_X$ that I carried out just after the proof of PBW. ∎

Thus

$$\begin{aligned} [x,[x,y]] &= x \otimes x \otimes y - 2x \otimes y \otimes x + y \otimes x \otimes x \\ [[x,y],y] &= x \otimes y \otimes y - 2y \otimes x \otimes y + y \otimes y \otimes x\,. \end{aligned}$$

9. Bracket computations

Let $\mathcal{H}$ be the image of H in $\mathcal{L}_X$. We are on our way to proving that $\mathcal{H}$ is a basis of $\mathcal{L}_X$. One important step is to show that the $\mathbb{Z}$-linear combinations of elements of $\mathcal{H}$ are closed under the bracket operation, or equivalently that if p and q are in $\mathcal{H}$ then $[p,q]$ is a $\mathbb{Z}$-linear combination of elements in $\mathcal{H}$. We shall in fact prove a stronger version of this claim suitable for a recursive proof.

9.1. Proposition. *If p and q are in H then $[p,q]$ is a $\mathbb{Z}$-linear combination of terms $[r,s]$ in $\mathcal{H}$ with $|r|+|s| = |p|+|q|$ and $r \ge \inf(p,q)$.*

The monomials in the sum are necessarily compound, since $|p|+|q| > 1$. The proof will be the basis of an explicit and practical algorithm.

Proof. By recursion on the set Σ of pairs (p,q) with p and q in H. This set Σ will be assigned a weak order:

$(r,s) \prec (p,q)$ *means that either (a)* $|r|+|s| < |p|+|q|$ *or (b)* $|r|+|s| = |p|+|q|$ *but* $\inf(r,s) > \inf(p,q)$.

Induction on $(\Sigma, \prec)$ is permissible, since the number of pairs less than a given one is finite.

Step 1. The minimal pairs in Σ are the (x,y) with x, y in X. If $x < y$ the monomial $\lfloor x,y \rfloor$ is in H; for (x,x) the bracket $[x,x]=0$; and for $x > y$ we have $[x,y] = -[y,x]$ with yx in H. So for these minimal elements of Σ the Proposition is true.

Step 2. Suppose now given (p,q) with $|p|+|q| > 2$. If $p > q$ we may replace it by $-[q,p]$. If $p = q$ then $[p,q] = 0$. So we may assume $p < q$, hence

$$\inf(p,q) = p\,.$$

Step 3. If q is in X, then $\lfloor p,q \rfloor$ is in H. Since $p \ge p$ we are done.

Step 4. So we are reduced to the case q is compound, say $q = \lfloor q_0, q_1 \rfloor$. If $q_0 \le p$ then $\lfloor p,q \rfloor$ itself is in H, and we are again done.

Step 5. Otherwise, we find ourselves with

$$p < q_0 < q_1\,,$$

with all three in H. This is the interesting situation. By Jacobi's identity

$$[p,[q_0,q_1]] = [[p,q_0],q_1] + [q_0,[p,q_1]]\,.$$

By recursion, since $|p|+|q_0| < |p|+|q|$, we can write

$$[p,q_0] = \sum n_{r,s}[r,s], \quad r \ge p = \inf(p,q_0), \quad |r|+|s| = |p|+|q_0|\,.$$

Each term in the sum is necessarily of degree > 1, hence compound. But then

$$[[p,q_0],q_1] = \sum n_{r,s}[[r,s],q_1]\,.$$

Here $|[r,s]|+|q_1| = |p|+|q|$ but—because of the Michel-Viennot condition—

$$\lfloor r,s \rfloor > r \ge p, q_1 > p$$

so that

$$\inf(\lfloor r,s \rfloor, q_1) > \inf(p,q) = p$$

and we can again apply induction to get

$$[[r,s],q_1] = \sum n_{r_*,s_*}[r_*,s_*], \quad r_* \ge \inf(\lfloor r,s \rfloor, q_1) > p$$

so we're done with the term $[[p, q_0], q_1]$.

Dealing with $[q_0, [p, q_1]]$ is similar.

This proof is almost identical to the one in [Serre:1965], but he imposes, as I have already mentioned, a stronger condition on the ordering of Hall sets.

Sample calculation. Impose lexicographic order on Hall monomials. The monomial $Y = [[x, y], [[x, y], y]]$ is a Hall monomial. How can we express

$$[x, Y] = [x, [[x, y], [[x, y], y]]]\,?$$

as a sum of Hall monomials? This bracket is not itself a Hall monomial. So we apply the Jacobi identity to get

$$\begin{aligned}
[x, [[x, y], [[x, y], y]]] &= {}^{(\boldsymbol{a})}\ [[x, [x, y]], [[x, y], y]] + {}^{(\boldsymbol{b})}\ [[x, y], {}^{(\boldsymbol{c})}\ [x, [[x, y], y]]] \\
{}^{(\boldsymbol{a})}\ [[x, [x, y]], [[x, y], y]] &= [[[x, [x, y]], [x, y]], y] + [[x, y], [[x, [x, y]], y]] \\
&= [[[x, [x, y]], [x, y]], y] - [[[x, [x, y]], y], [x, y]] \\
{}^{(\boldsymbol{c})}\ [x, [[x, y], y]] &= [[x, [x, y]], y] + [[x, y], [x, y]] \\
&= [[x, [x, y]], y] \\
{}^{(\boldsymbol{b})}\ [[x, y],\ [x, [[x, y], y]]] &= \quad [[x, y], [[x, [x, y]], y]] \\
&= -[[x, [[x, y], y]], [x, y]] \\
[x, [[x, y], [[x, y], y]]] &= [[[x, [x, y]], [x, y]], y] - 2\, [[[x, [x, y]], y], [x, y]]\,.
\end{aligned}$$

We can now finally prove

9.2. Theorem. *A Hall set is a basis of $U(\mathcal{L}_X)$.*

Proof. The previous result shows that a Hall set contains X and linear combinations are closed under bracket operations. So it spans $\mathcal{L}_X$. But Corollary 7.6 shows that the elements of a Hall set remain linearly independent when embedded in $\bigotimes^\bullet R_X$, hence certainly in $\mathcal{L}_X$.

We know that $\bigotimes^\bullet R_X$ is $U(\mathcal{L}_X)$ and we now know that $\mathcal{H}$ is a basis of $\mathcal{L}_X$. We therefore also know by PBW that every element of $\bigotimes^\bullet R_X$ may be expressed as a linear combination of a product of weakly decreasing product of elements of Hall monomials. I have explained very briefly how to find such an expression in a special case, but let me recapitulate.

We start with $x_1 \otimes \ldots \otimes x_n$. At any step in this computation we are considering a tensor product $p_1 \otimes \ldots \otimes p_m$ of Hall monomials. We scan from left to right, If $p_1 \geq \ldots \geq p_m$ then there is nothing to be done. Otherwise, we shall have $p_1 \geq \ldots \geq p_k < p_{k+1}$ for some k. We replace this by

$$p_1 \otimes \ldots \otimes p_{k+1} \otimes p_k \otimes \ldots \otimes p_m + p_1 \otimes \ldots \otimes [p_k, p_{k+1}] \otimes \ldots \otimes p_m$$

According to Lemma 7.2, the sequence of monomials in each term is admissible, and in particular each one of the p_i is a Hall monomial.

10. Dynkin, Friedrichs, Specht, and Wever

How do we tell whether a given element of $\bigotimes^{\bullet} V$ is a Lie polynomial? There are two useful characterizations of the subspace L_X embedded in $\bigotimes^{\bullet} R_X$.

FIRST CHARACTERIZATION. This is due to [Friedrichs:1953].

Assume here that $\mathfrak{g}$ is an arbitrary Lie algebra. The direct sum of two Lie algebras is defined so that the two summands commute. It is therefore easy to see:

10.1. Proposition. *If $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{k}$ then the map $x \oplus y \mapsto x \otimes 1 + 1 \otimes y$ induces an isomorphism*

$$U(\mathfrak{g}) \cong U(\mathfrak{h}) \otimes U(\mathfrak{k})\,.$$

The diagonal map $\Delta\colon \mathfrak{g} \to \mathfrak{g} \oplus \mathfrak{g}$ induces a ring homomorphism

$$\Delta\colon\; U(\mathfrak{g}) \longrightarrow U(\mathfrak{g}) \otimes U(\mathfrak{g})\,.$$

A **primitive element** of $U(\mathfrak{g})$ is an x such that

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

10.2. Proposition. *The image of $\mathfrak{g}$ in $U(\mathfrak{g})$ is exactly the set of primitive elements.*

Proof. It is clear that $\Delta x = x \otimes 1 + 1 \otimes x$ for x in $\mathfrak{g}$.

For the converse, suppose at first that $\mathfrak{g}$ is abelian, say of dimension r. The universal enveloping algebra U may be identified with the polynomial algebra in r variables x_i, the algebra $U \otimes U$ with the polynomial algebra in $2r$ variables x_i, y_i. The diagonal embedding takes x_i to $x_i + y_i$ and $P(x_i)$ to $P(x_i + y_i)$. Thus $P(x) \otimes 1$ may be identified with $P(x_i)$ and $1 \otimes P(x)$ may be identified with $P(y_i)$ and a primitive element is a polynomial P with $P(x_i + y_i) = P(x_i) + P(y_i)$. So the following Lemma concludes this case:

10.3. Lemma. *Suppose P is a polynomial such that $P(x+y) = P(x) + P(y)$ identically. Then $P(x) = ax$ for some constant a.*

Proof. This is elementary, but I'll still give details. If P is any polynomial, recall the difference functions $[\Delta^k P](n)$ defined by recursion

$$\Delta^0 P = P, \quad [\Delta^k P](n) = [\Delta^{k-1} P](n+1) - [\Delta^{k-1} P](n)\,.$$

The degree of $\Delta^k P$ is k less than that of P, and if $P = \sum p_i x^i$ has degree n then $\Delta^n P = n! c_n$. In our case, $P(n) = nP(1)$, so $\Delta^k P = 0$ for $k > 1$. This implies that P must have degree at most one. Since $P(0) = 0$ it is linear. ∎

The general case reduces to the abelian case by looking at the graded enveloping algebras. ∎

SECOND CHARACTERIZATION. This is due independently and almost simultaneously to [Dynkin:1947], [Specht:1948], and [Wever:1947].

It amounts to an extremely explicit idempotent projection Ω from $\bigotimes^{\bullet} R_X$ to $\mathbb{Q} \otimes L_X$. Define it inductively by the specifications

$$\begin{aligned} \Omega(1) &= 0 \\ \Omega(x) &= x \\ \Omega(x \otimes p) &= [x, \Omega(p)]\,. \end{aligned}$$

For example, Ω takes

$$\begin{aligned} x_1 &\longmapsto x_1 \\ x_1 \otimes x_2 &\longmapsto [x_1, x_2] \\ x_1 \otimes x_2 \otimes x_3 &\longmapsto [x_1, [x_2, x_3]] \\ x_1 \otimes x_2 \otimes x_3 \otimes x_4 &\longmapsto [x_1, [x_2, [x_3, x_4]]] \\ &\dots \end{aligned}$$

I'll call an element of L_X in the same form as one of these a **right-normal** element. The formal definition is by induction: (a) x is right-normal if x is in X; (b) $[x, p]$ is right-normal if p is.

10.4. Lemma. *The space L_X is spanned by right-normal elements.*

In other words, the map Ω is surjective.

Proof. It must be shown that if p, q are monomials in L_X then $[p, q]$ can be expressed as a linear combination of right-normal elements. The proof goes by induction on the degree of p.

- If $p = x$ has degree one, then $[x, y]$ is right-normal, and if q is right-normal then so is $[x, q]$. So one may apply induction on the degree of q.

The induction assumption now is that $[p, q]$ is a sum of right-normal elements if p and q are Lie polynomials, with the degree of p less than n.

- Say p has degree n. We may assume $p = [p_1, p_2]$ in which the degree of each p_i is less than n. But by Jacobi's identity

$$[p, q] = [[p_1, p_2], q] = -[[p_2, q], p_1] - [[q, p_1], p_2] \, .$$

To which we may apply induction. ∎

Because $\bigotimes^\bullet R_X$ may be identified with $U(\mathcal{L}_X)$, the map $x \mapsto \mathrm{ad}_x$ may be extended uniquely to a ring homomorphism θ from $\bigotimes^\bullet R_X$ to $\mathrm{End}(L_X)$:

$$x_1 \otimes \ldots \otimes x_n \longmapsto \mathrm{ad}_{x_1} \ldots \mathrm{ad}_{x_n} \, .$$

The following is the characteristic property of Ω:

10.5. Lemma. *For x in $\bigotimes^\bullet R_X$ and y in L_X*

$$\Omega(x \otimes y) = \theta(x)\Omega(y) \, .$$

Proof. This may be proved by induction on the degree of x, but for y in right-normal form it is an immediate consequence of the definition of Ω. Apply the Lemma. ∎

10.6. Theorem. *Suppose p to be an element of $\bigotimes^n R_X$. Then it is a Lie polynomial if and only if $\Omega(p) = np$.*

Proof. There are many proofs in the literature, of varying illumination. I follow the treatment in §4.8 of [Serre:1965].

Proof of Theorem 10.6. If $\Omega(x) = nx$, then nx lies in L_X since all of the image lies in L_X. But L_X is free in $\bigotimes^\bullet R_X$, so x also lie in L_X.

Now to show that if x lies in $L_X^{(n)}$ then $\Omega(x) = nx$. We may assume that $x = [p, q]$ with degrees of p, q equal to m, $n - m$ less than n. But then by the Lemma, since $\theta(x)$ extends ad_x:

$$\begin{aligned} \Omega([p, q]) &= \Omega(pq - qp) = \theta(p)\Omega(q) - \theta(q)\Omega(p) \\ &= m\theta(p)q - (n - m)\theta(q)p \\ &= m[p, q] - (n - m)[q, p] = n[p, q] \, . \end{aligned}$$

∎

11. Baker, Campbell, Hausdorff, and Dynkin

We can extend L_X (the copy of $\mathcal{L}_X$ in the tensor algebra) to a formal power series ring, by completing the tensor algebra with respect to terms of degree greater than 0. Let $\mathfrak{m}$ be the ideal of such terms, and $\widehat{\mathfrak{m}}$ be its completion . In this context, an infinite series makes sense. For example, if x lies in $\mathfrak{m}$ then the inverse of $1+x$ is $1-x+x^2-x^3+x^4-\cdots$. We are dealing here with associative but non-commutative formal power series.

We can define series $\exp$ and $\log$ in the natural way:

$$\begin{aligned}\exp(x) &= 1+x+\frac{x^2}{2!}+\cdots \\ \log(1+x) &= x-\frac{x^2}{2}+\frac{x^3}{3}-\cdots\end{aligned}$$

for x in $\mathfrak{m}$. They are inverse to each other as formal series, so are inverse here, too, and induce isomorphisms:

$$\widehat{\mathfrak{m}} \cong 1+\widehat{\mathfrak{m}}\,.$$

Formal multiplication of series makes a group out of the right and side:

$$(1+a_1+a_2+a_3+\cdots)(1+b_1+b_2+b_3+\cdots) = \big(1+(a_1+b_1)+(a_2+a_1b_1+b_2)+(a_3+a_2b_1+a_1b_2)+b_3)\cdots\big)$$

since the k-term lies in $\widehat{\mathfrak{m}}^k$. The rough idea of the next result is that the structure of this group depends in a natural way entirely on the structure of the Lie algebra $\mathcal{L}_X \cong L_X$. When $R=\mathbb{R}$ the formal series defined here correspond to convergent series, and this result has as consequence that the local structure of a Lie group depends only on the structure of its Lie algebra.

Let $X=\{x,y\}$. Define z in the tensor product algebra $\bigotimes^\bullet R_X$ of R_X by the formula

$$z=\log\left(e^x e^y\right).$$

Explicitly, we have

$$e^x e^y = \left(1+x+\frac{x^2}{2}+\frac{x^3}{3!}+\cdots\right)\left(1+y+\frac{y^2}{2}+\frac{y^3}{3!}+\cdots\right) = \sum_{p,q=0}^{\infty}\frac{x^p y^q}{p!\,q!}\,.$$

and then

$$\begin{aligned}\log\left(e^x e^y\right) &= \sum_{r=1}\frac{(-1)^{r+1}}{r}\left(\sum_{p+q\ge 1}^{\infty}\frac{x^p y^q}{p!\,q!}\right)^r \\ &= \sum_m\sum_{p+q\ge 1}\frac{(-1)^{m+1}}{m}\,\frac{x^{p_1}y^{q_1}\ldots x^{p_m}y^{q_m}}{p_1!\,q_1!\ldots p_m!\,q_m!}\,.\end{aligned}$$

The first few terms in the expansion of z can be found without too much trouble:

$$\begin{aligned}&x+y \\ &\frac{1}{2}[x,y] \\ &\frac{1}{12}\big([x,[x,y]]+[y,[y,x]]\big)\end{aligned}$$

This suggests:

11.1. Proposition. *The series $z=\log\big(\exp(x)\exp(y)\big)$ lies in the Lie algebra $\widehat{\mathcal{L}}_X$.*

That is to say, its homogeneous terms are Lie polynomials.

Proof. We apply Proposition 10.2, after first proving:

11.2. Lemma. *The exponential map is an isomorphism of primitive z in $\widehat{\mathfrak{m}}$ with z in $1 + \widehat{\mathfrak{m}}$ such that $\Delta(z) = \Delta(z) \otimes \Delta(z)$.*

It has to be shown only that z is primitive in the tensor algebra. But we can see easily that

$$\Delta(e^z) = e^{\Delta}(z) = e^{z\otimes 1 + 1\otimes z} = e^z \otimes e^z \,. \;∎$$

It would be nice to have an algorithmic proof of Proposition 11.1, one that at least outlined an algorithm to show explicitly the chain of steps implicitly involved in the theorem. The proof in [Knapp:2002] comes close.

There is an explicit expression for z, first found rather late in the history of the subject—in [Dynkin:1947]—although it is implicit in earlier work of Baker. It is on the one hand not as important as Lemma 11.2, but on the other it is not so difficult to prove once the first half is known, and rather satisfying.

11.3. Theorem. *For x and y in $\mathfrak{g}$ set*

$$\log(e^x e^y) = \sum H_n$$

where H_n is the term of homogeneous degree n. Then

$$H_n = \frac{1}{n} \sum_{p+q=n} \left(H^{(1)}_{p,q} + H^{(2)}_{p,q} \right)$$

where

$$H^{(1)}_{p,q} = \sum \frac{(-1)^{m+1}}{m} \frac{\mathrm{ad}_x^{p_1} \ldots \mathrm{ad}_x^{p_m}(y)}{p_1!q_1!\ldots p_m!} \quad \textit{with the sum over } \begin{array}{l} p_1 + \cdots p_m = p \\ q_1 + \cdots + q_{m-1} = q-1 \\ p_i + q_i \ge 1 \\ p_m \ge 1 \end{array}$$

$$H^{(2)}_{p,q} = \sum \frac{(-1)^{m+1}}{m} \frac{\mathrm{ad}_x^{p_1} \ldots \mathrm{ad}_x^{q_{m-1}}(x)}{p_1!q_1!\ldots q_{m-1}!} \quad \textit{with the sum over } \begin{array}{l} p_1 + \cdots p_{m-1} = p-1 \\ q_1 + \cdots + q_{m-1} = q \\ p_i + q_i \ge 1 \end{array}$$

$(, \; , \; p_i + q_i \ge 1)$

Proof. Show that each term is the projection under Ω of the term in the explicit series. ∎

12. References

1. Peter Bendix and Donald Knuth, 'Simple word problems in universal algebras', in **Computational problems in abstract algebra**, edited by John Leech, 263–297, 1970.

2. George Bergman,'The Diamond Lemma for ring theory', *Advances in Mathematics* **29** (1978), 178–218.

3. G. Birkhoff, 'Representability of Lie algebras and Lie groups by matrices', *Annals of Mathematics* **38** (1937), 526–532.

4. Evgeny Dynkin, 'Calculation of the coefficients in the Campbell-Hausdorff formula', pages 31–36 in **Selected papers of E. B. Dynkin**, American Mathematical Society, 2000. Originally published (in Russian) in *Doklady Akad. Nauk SSSR* **57** (1947), 323–326.

5. Kurt Friedrichs, 'Mathematical aspects of the quantum theory of fields V', *Communications in Pure and Applied Mathematics* **6** (1953), 1–72.

6. Willem de Graaf, 'Hall and Gröbner bases and rewriting in free Lie algebras (extended abstract)', 1999. Available at

`http://citeseer.ist.psu.edu/330216.html`

7. Marshall Hall, 'A basis for free Lie algebras and higher commutators in free groups', *Proceedings of the American Mathematical Society* **1** (1950), 575–81.

8. Philip Hall, 'A contribution to the theory of groups of prime power order', *Proceedings of the London Mathematical Society* **36** (1934), 29–95.

9. G. H. Hardy and E. Wright, **An introduction to the theory of numbers**, Oxford, 1960.

10. Anthony Knapp, **Beyond an introduction**, Birkhäuser, 2002.

11. Wilhelm Magnus, 'Über beziehung zwischen höher Kommutatoren', *Journal für die reine und angewandte Mathematik* **177** (1937), 105–115.

12. M. H. A. Newman, 'On theories with a combinatorial definition of "equivalence"', *Annals of Mathematics* **43** (1942), 223–243.

13. Christophe Reutenauer, **Free Lie algebras**, Oxford University Press, 1993.

14. Wulf Rossmann, **Lie groups**, Oxford University Press, 2002.

15. Jean-Pierre Serre, **Lie algebras and Lie groups**, Benjamin, 1965.

16. W. Specht, 'Die linearen Beziehungen zwischen höheren Kommutatoren', *Mathematische Zeitschrift* **51** (1948), 367–376.

17. Gérard Viennot, **Algèbres de Lie libres et monoïdes libres**, *Lecture Notes in Mathematics* **691**, Springer, 1978.

18. Wever, 'Überr invarianten in Lie'schen Ringen', *Mathematische Annalen* **120** (1947), 563-580.