Lie Algebras

Shlomo Sternberg

February 5, 2004

Chapter 1

The Campbell-Baker-Hausdorff Formula

1.1 The problem.

Recall the power series:

$$\exp X = 1 + X + \frac{1}{2}X^2 + \frac{1}{3!}X^3 + \cdots, \quad \log(1+X) = X - \frac{1}{2}X^2 + \frac{1}{3}X^3 + \cdots.$$

We want to study these series in a ring where convergence makes sense; for example in the ring of $n \times n$ matrices. The exponential series converges everywhere, and the series for the logarithm converges in a small enough neighborhood of the origin. Of course,

$$\log(\exp X) = X; \quad \exp(\log(1+X)) = 1 + X$$

where these series converge, or as formal power series.

In particular, if A and B are two elements which are close enough to 0 we can study the convergent series

$$\log[(\exp A)(\exp B)]$$

which will yield an element C such that $\exp C = (\exp A)(\exp B)$. The problem is that A and B need not commute. For example, if we retain only the linear and constant terms in the series we find

$$\log[(1 + A + \cdots)(1 + B + \cdots)] = \log(1 + A + B + \cdots) = A + B + \cdots.$$

On the other hand, if we go out to terms second order, the non-commutativity begins to enter:

$$\log[(1 + A + \frac{1}{2}A^2 + \cdots)(1 + B + \frac{1}{2}B^2 + \cdots)] =$$

$$\begin{aligned}A+B+\frac{1}{2}A^2+AB+\frac{1}{2}B^2-\frac{1}{2}(A+B+\cdots)^2\\ =A+B+\frac{1}{2}[A,B]+\cdots\end{aligned}$$

where

$$[A,B]:=AB-BA \tag{1.1}$$

is the **commutator** of A and B, also known as the **Lie bracket** of A and B.

Collecting the terms of degree three we get, after some computation,

$$\frac{1}{12}\left(A^2B+AB^2+B^2A+BA^2-2ABA-2BAB]\right)=\frac{1}{12}[A,[A,B]]+\frac{1}{12}[B,[B,A]].$$

This suggests that the series for $\log[(\exp A)(\exp B)]$ can be expressed entirely in terms of successive Lie brackets of A and B. This is so, and is the content of the Campbell-Baker-Hausdorff formula.

One of the important consequences of the mere existence of this formula is the following. Suppose that $\mathbf{g}$ is the Lie algebra of a Lie group G. Then the *local* structure of G near the identity, i.e. the rule for the product of two elements of G sufficiently closed to the identity is determined by its Lie algebra $\mathbf{g}$. Indeed, the exponential map is locally a diffeomorphism from a neighborhood of the origin in $\mathbf{g}$ onto a neighborhood W of the identity, and if $U\subset W$ is a (possibly smaller) neighborhood of the identity such that $U\cdot U\subset W$, the the product of $a=\exp\xi$ and $b=\exp\eta$, with $a\in U$ and $b\in U$ is then completely expressed in terms of successive Lie brackets of ξ and η.

We will give two proofs of this important theorem. One will be geometric - the explicit formula for the series for $\log[(\exp A)(\exp B)]$ will involve integration, and so makes sense over the real or complex numbers. We will derive the formula from the "Maurer-Cartan equations" which we will explain in the course of our discussion. Our second version will be more algebraic. It will involve such ideas as the universal enveloping algebra, comultiplication and the Poincaré-Birkhoff-Witt theorem. In both proofs, many of the key ideas are at least as important as the theorem itself.

1.2 The geometric version of the CBH formula.

To state this formula we introduce some notation. Let ad A denote the operation of bracketing on the left by A, so

$$\mathrm{ad}A(B):=[A,B].$$

Define the function ψ by

$$\psi(z)=\frac{z\log z}{z-1}$$

which is defined as a convergent power series around the point $z=1$ so

$$\psi(1+u)=(1+u)\frac{\log(1+u)}{u}=(1+u)(1-\frac{u}{2}+\frac{u^2}{3}+\cdots)=1+\frac{u}{2}-\frac{u^2}{6}+\cdots.$$

In fact, we will also take this as a *definition* of the formal power series for ψ in terms of u. The Campbell-Baker-Hausdorff formula says that

$$\log((\exp A)(\exp B)) = A + \int_0^1 \psi\left((\exp \mathrm{ad}\ A)(\exp t \mathrm{ad}\ B)\right) B dt. \tag{1.2}$$

Remarks.

1. The formula says that we are to substitute

$$u = (\exp \mathrm{ad}\ A)(\exp t\mathrm{ad}\ B) - 1$$

into the definition of ψ, apply this operator to the element B and then integrate. In carrying out this computation we can ignore all terms in the expansion of ψ in terms of ad A and ad B where a factor of ad B occurs on the right, since $(\mathrm{ad}\ B)B = 0$. For example, to obtain the expansion through terms of degree three in the Campbell-Baker-Hausdorff formula, we need only retain quadratic and lower order terms in u, and so

$$\begin{aligned} u &= \mathrm{ad}\ A + \frac{1}{2}(\mathrm{ad}\ A)^2 + t\mathrm{ad}\ B + \frac{t^2}{2}(\mathrm{ad}\ B)^2 + \cdots \\ u^2 &= (\mathrm{ad}\ A)^2 + t(\mathrm{ad}\ B)(\mathrm{ad}\ A) + \cdots \\ \int_0^1 \left(1 + \frac{u}{2} - \frac{u^2}{6}\right) dt &= 1 + \frac{1}{2}\mathrm{ad}\ A + \frac{1}{12}(\mathrm{ad}\ A)^2 - \frac{1}{12}(\mathrm{ad}\ B)(\mathrm{ad}\ A) + \cdots, \end{aligned}$$

where the dots indicate either higher order terms or terms with ad B occurring on the right. So up through degree three (1.2) gives

$$\log(\exp A)(\exp B) = A + B + \frac{1}{2}[A,B] + \frac{1}{12}[A,[A,B]] - \frac{1}{12}[B,[A,B]] + \cdots$$

agreeing with our preceding computation.

2. The meaning of the exponential function on the left hand side of the Campbell-Baker-Hausdorff formula differs from its meaning on the right. On the right hand side, exponentiation takes place in the algebra of endomorphisms of the ring in question. In fact, we will want to make a fundamental reinterpretation of the formula. We want to think of A, B, etc. as elements of a Lie algebra, $\mathbf{g}$. Then the exponentiations on the right hand side of (1.2) are still taking place in $\mathrm{End}(\mathbf{g})$. On the other hand, if $\mathbf{g}$ is the Lie algebra of a Lie group G, then there is an exponential map: exp: $\mathbf{g} \to G$, and this is what is meant by the exponentials on the left of (1.2). This exponential map is a diffeomorphism on some neighborhood of the origin in $\mathbf{g}$, and its inverse, log, is defined in some neighborhood of the identity in G. This is the meaning we will attach to the logarithm occurring on the left in (1.2).

3. The most crucial consequence of the Campbell-Baker-Hausdorff formula is that it shows that the local structure of the Lie group G (the multiplication law for elements near the identity) is completely determined by its Lie algebra.

4. For example, we see from the right hand side of (1.2) that group multiplication and group inverse are analytic if we use exponential coordinates.

5. Consider the function τ defined by

$$\tau(w) := \frac{w}{1-e^{-w}}. \tag{1.3}$$

This is a familiar function from analysis, as it enters into the Euler-Maclaurin formula, see below. (It is the exponential generating function of $(-1)^k b_k$ where the b_k are the Bernoulli numbers.) Then

$$\psi(z) = \tau(\log z).$$

6. The formula is named after three mathematicians, Campbell, Baker, and Hausdorff. But this is a misnomer. Substantially earlier than the works of any of these three, there appeared a paper by Friedrich Schur, "Neue Begruendung der Theorie der endlichen Transformationsgruppen," *Mathematische Annalen* **35** (1890), 161-197. Schur writes down, as convergent power series, the composition law for a Lie group in terms of "canonical coordinates", i.e., in terms of linear coordinates on the Lie algebra. He writes down recursive relations for the coefficients, obtaining a version of the formulas we will give below. I am indebted to Prof. Schmid for this reference.

Our strategy for the proof of (1.2) will be to prove a differential version of it:

$$\frac{d}{dt}\log\left((\exp A)(\exp tB)\right) = \psi\left((\exp \mathrm{ad}\ A)(\exp t\ \mathrm{ad}\ B)\right)B. \tag{1.4}$$

Since $\log\exp A = A$ when $t = 0$, integrating (1.4) from 0 to 1 will prove (1.2). Let us define $\Gamma = \Gamma(t) = \Gamma(t, A, B)$ by

$$\Gamma = \log\left((\exp A)(\exp tB)\right). \tag{1.5}$$

Then

$$\exp\Gamma = \exp A \exp tB$$

and so

$$\begin{aligned} \frac{d}{dt}\exp\Gamma(t) &= \exp A\frac{d}{dt}\exp tB \\ &= \exp A(\exp tB)B \\ &= (\exp\Gamma(t))B \quad \text{so} \\ (\exp -\Gamma(t))\frac{d}{dt}\exp\Gamma(t) &= B. \end{aligned}$$

We will prove (1.4) by finding a general expression for

$$\exp(-C(t))\frac{d}{dt}\exp(C(t))$$

where $C = C(t)$ is a curve in the Lie algebra, $\mathbf{g}$, see (1.11) below.

In our derivation of (1.4) from (1.11) we will make use of an important property of the adjoint representation which we might as well state now: For any $g \in G$, define the linear transformation

$$\mathrm{Ad}\ g : \mathbf{g} \to \mathbf{g} : X \mapsto gXg^{-1}.$$

(In geometrical terms, this can be thought of as follows: (The differential of) Left multiplication by g carries $\mathbf{g} = T_I(G)$ into the tangent space, $T_g(G)$ to G at the point g. Right multiplication by g^{-1} carries this tangent space back to $\mathbf{g}$ and so the combined operation is a linear map of $\mathbf{g}$ into itself which we call Ad g. Notice that Ad is a representation in the sense that

$$\mathrm{Ad}\ (gh) = (\mathrm{Ad}\ g)(\mathrm{Ad}\ h)\ \ \forall g, h \in G.$$

In particular, for any $A \in \mathbf{g}$, we have the one parameter family of linear transformations $\mathrm{Ad}(\exp tA)$ and

$$\begin{aligned} \frac{d}{dt}\mathrm{Ad}\ (\exp tA)X &= (\exp tA)AX(\exp -tA) + (\exp tA)X(-A)(\exp -tA) \\ &= (\exp tA)[A, X](\exp -tA) \ \ \text{so} \\ \frac{d}{dt}\mathrm{Ad}\ \exp tA &= \mathrm{Ad}(\exp tA) \circ \mathrm{ad}\ A. \end{aligned}$$

But ad A is a linear transformation acting on $\mathbf{g}$ and the solution to the differential equation

$$\frac{d}{dt}M(t) = M(t)\mathrm{ad}\ A, \quad M(0) = I$$

(in the space of linear transformations of $\mathbf{g}$) is $\exp t\,\mathrm{ad}\,A$. Thus $\mathrm{Ad}(\exp tA) = \exp(t\,\mathrm{ad}\,A)$. Setting $t = 1$ gives the important formula

$$\mathrm{Ad}\ (\exp A) = \exp(\mathrm{ad}\ A). \tag{1.6}$$

As an application, consider the Γ introduced above. We have

$$\begin{aligned} \exp(\mathrm{ad}\ \Gamma) &= \mathrm{Ad}\ (\exp \Gamma) \\ &= \mathrm{Ad}\ ((\exp A)(\exp tB)) \\ &= (\mathrm{Ad}\ \exp A)(\mathrm{Ad}\ \exp tB) \\ &= (\exp \mathrm{ad}\ A)(\exp \mathrm{ad}\ tB) \end{aligned}$$

hence

$$\mathrm{ad}\ \Gamma = \log((\exp \mathrm{ad}\ A)(\exp \mathrm{ad}\ tB)). \tag{1.7}$$

1.3 The Maurer-Cartan equations.

If G is a Lie group and $\gamma = \gamma(t)$ is a curve on G with $\gamma(0) = A \in G$, then $A^{-1}\gamma$ is a curve which passes through the identity at $t = 0$. Hence $A^{-1}\gamma'(0)$ is a tangent vector at the identity, i.e. an element of $\mathbf{g}$, the Lie algebra of G.

In this way, we have defined a linear differential form θ on G with values in $\mathbf{g}$. In case G is a subgroup of the group of all invertible $n \times n$ matrices (say over the real numbers), we can write this form as

$$\theta = A^{-1}dA.$$

We can then think of the A occurring above as a collection of n^2 real valued functions on G (the matrix entries considered as functions on the group) and dA as the matrix of differentials of these functions. The above equation giving θ is then just matrix multiplication. For simplicity, we will work in this case, although the main theorem, equation (1.8) below, works for any Lie group and is quite standard.

The definitions of the groups we are considering amount to constraints on A, and then differentiating these constraints show that $A^{-1}dA$ takes values in $\mathbf{g}$, and gives a description of $\mathbf{g}$. It is best to explain this by examples:

- O(n): $AA^{\dagger} = I, \quad dAA^{\dagger} + AdA^{\dagger} = 0$ or

  $$A^{-1}dA + \left(A^{-1}dA\right)^{\dagger} = 0.$$

  $o(n)$ consists of antisymmetric matrices.

- Sp(n): Let

  $$J := \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

  and let Sp(n) consist of all matrices satisfying

  $$AJA^{\dagger} = J.$$

  Then

  $$dAJa^{\dagger} + AJdA^{\dagger} = 0$$

  or

  $$(A^{-1}dA)J + J(A^{-1}dA)^{\dagger} = 0.$$

  The equation $BJ + JB^{\dagger} = 0$ defines the Lie algebra sp(n).

- Let J be as above and define Gl(n,C) to consist of all invertible matrices satisfying

  $$AJ = JA.$$

  Then

  $$dAJ = JdA = 0.$$

  and so

  $$A^{-1}dAJ = A^{-1}JdA = JA^{-1}dA.$$

We return to general considerations: Let us take the exterior derivative of the defining equation $\theta = A^{-1}dA$. For this we need to compute $d(A^{-1})$: Since

$$d(AA^{-1}) = 0$$

we have

$$dA \cdot A^{-1} + Ad(A^{-1}) = 0$$

or

$$d(A^{-1}) = -A^{-1}dA \cdot A^{-1}.$$

This is the generalization to matrices of the formula in elementary calculus for the derivative of $1/x$. Using this formula we get

$$d\theta = d(A^{-1}dA) = -(A^{-1}dA \cdot A^{-1}) \wedge dA = -A^{-1}dA \wedge A^{-1}dA$$

or the **Maurer-Cartan equation**

$$d\theta + \theta \wedge \theta = 0. \tag{1.8}$$

If we use commutator instead of multiplication we would write this as

$$d\theta + \frac{1}{2}[\theta, \theta] = 0. \tag{1.9}$$

The Maurer-Cartan equation is of central importance in geometry and physics, far more important than the Campbell-Baker-Hausdorff formula itself.

Suppose we have a map $g : \mathbf{R}^2 \to G$, with s, t coordinates on the plane. Pull θ back to the plane, so

$$g^*\theta = g^{-1}\frac{\partial g}{\partial s}ds + g^{-1}\frac{\partial g}{\partial t}dt$$

Define

$$\alpha = \alpha(s, t) := g^{-1}\frac{\partial g}{\partial s}$$

and

$$\beta := \beta(s, t) = g^{-1}\frac{\partial g}{\partial t}$$

so that

$$g^*\theta = \alpha ds + \beta dt.$$

Then collecting the coefficient of $ds \wedge dt$ in the Maurer Cartan equation gives

$$\frac{\partial \beta}{\partial s} - \frac{\partial \alpha}{\partial t} + [\alpha, \beta] = 0. \tag{1.10}$$

This is the version of the Maurer Cartan equation we shall use in our proof of the Campbell Baker Hausdorff formula. Of course this version is completely equivalent to the general version, since a two form is determined by its restriction to all two dimensional surfaces.

1.4 Proof of CBH from Maurer-Cartan.

Let $C(t)$ be a curve in the Lie algebra $\mathbf{g}$ and let us apply (1.10) to

$$g(s,t) := \exp[sC(t)]$$

so that

$$\begin{aligned}
\alpha(s,t) &= g^{-1}\frac{\partial g}{\partial s} \\
&= \exp[-sC(t)]\exp[sC(t)]C(t) \\
&= C(t) \\
\beta(s,t) &= g^{-1}\frac{\partial g}{\partial t} \\
&= \exp[-sC(t)]\frac{\partial}{\partial t}\exp[sC(t)] \text{ so by (1.10)} \\
\frac{\partial\beta}{\partial s} - C'(t) + [C(t),\beta] &= 0.
\end{aligned}$$

For fixed t consider the last equation as the differential equation (in s)

$$\frac{d\beta}{ds} = -(\mathrm{ad}\ C)\beta + C', \quad \beta(0) = 0$$

where $C := C(t),\ C' := C'(t)$.

If we expand $\beta(s,t)$ as a formal power series in s (for fixed t):

$$\beta(s,t) = a_1 s + a_2 s^2 + a_3 s^3 + \cdots$$

and compare coefficients in the differential equation we obtain $a_1 = C'$, and

$$na_n = -(\mathrm{ad}\ C)a_{n-1}$$

or

$$\beta(s,t) = sC'(t) + \frac{1}{2}s(-\mathrm{ad}\ C(t))C'(t) + \cdots + \frac{1}{n!}s^n(-\mathrm{ad}\ C(t))^{n-1}C'(t) + \cdots .$$

If we define

$$\phi(z) := \frac{e^z - 1}{z} = 1 + \frac{1}{2!}z + \frac{1}{3!}z^2 + \cdots$$

and set $s = 1$ in the expression we derived above for $\beta(s,t)$ we get

$$\exp(-C(t))\frac{d}{dt}\exp(C(t)) = \phi(-\mathrm{ad}\ C(t))C'(t). \tag{1.11}$$

Now to the proof of the Campbell-Baker-Hausdorff formula. Suppose that A and B are chosen sufficiently near the origin so that

$$\Gamma = \Gamma(t) = \Gamma(t,A,B) := \log((\exp A)(\exp tB))$$

is defined for all $|t| \leq 1$. Then, as we remarked,

$$\exp \Gamma = \exp A \exp tB$$

so $\exp \operatorname{ad} \Gamma = (\exp \operatorname{ad}\ A)(\exp t\ \operatorname{ad}\ B)$ and hence

$$\operatorname{ad}\ \Gamma = \log\left((\exp \operatorname{ad}\ A)(\exp t\ \operatorname{ad}\ B)\right).$$

We have

$$\begin{aligned}
\frac{d}{dt}\exp \Gamma(t) &= \exp A \frac{d}{dt}\exp tB \\
&= \exp A(\exp tB)B \\
&= (\exp \Gamma(t)B \quad \text{so} \\
(\exp -\Gamma(t))\frac{d}{dt}\exp \Gamma(t) &= B \text{ and therefore} \\
\phi(-\operatorname{ad}\ \Gamma(t))\Gamma'(t) &= B \quad \text{by (1.11) so} \\
\phi(-\log((\exp \operatorname{ad}\ A)(\exp t\ \operatorname{ad}\ B)))\Gamma'(t) &= B.
\end{aligned}$$

Now for $|z-1| < 1$

$$\begin{aligned}
\phi(-\log z) &= \frac{e^{-\log z}-1}{-\log z} \\
&= \frac{z^{-1}-1}{-\log z} \\
&= \frac{z-1}{z\log z} \quad \text{so} \\
\psi(z)\phi(-\log z) &\equiv 1 \quad \text{where } \psi(z) := \frac{z\log z}{z-1} \text{ so} \\
\Gamma'(t) &= \psi\left((\exp \operatorname{ad}\ A)(\exp t\operatorname{ad}\ B)\right)B.
\end{aligned}$$

This proves (1.4) and integrating from 0 to 1 proves (1.2).

1.5 The differential of the exponential and its inverse.

Once again, equation (1.11), which we derived from the Maurer-Cartan equation, is of significant importance in its own right, perhaps more than the use we made of it - to prove the Campbell-Baker-Hausdorff theorem. We will rewrite this equation in terms of more familiar geometric operations, but first some preliminaries:

The exponential map exp sends the Lie algebra **g** into the corresponding Lie group, and is a differentiable map. If $\xi \in \mathbf{g}$ we can consider the differential of exp at the point ξ:

$$d(\exp)_\xi : \mathbf{g} = T\mathbf{g}_\xi \to TG_{\exp \xi}$$

where we have identified $\mathbf{g}$ with its tangent space at ξ which is possible since $\mathbf{g}$ is a vector space. In other words, $d(\exp)_\xi$ maps the tangent space to $\mathbf{g}$ at the point ξ into the tangent space to G at the point $\exp(\xi)$. At $\xi = 0$ we have

$$d(\exp)_0 = \mathrm{id}$$

and hence, by the implicit function theorem, $d(\exp)_\xi$ is invertible for sufficiently small ξ. Now the Maurer-Cartan form, evaluated at the point $\exp\xi$ sends $TG_{\exp\xi}$ back to $\mathbf{g}$:

$$\theta_\xi : TG_{\exp\xi} \to \mathbf{g}.$$

Hence

$$\theta_\xi \circ d(\exp)_\xi : \mathbf{g} \to \mathbf{g}$$

and is invertible for sufficiently small ξ. We claim that

$$\tau(\operatorname{ad}\xi) \circ \left(\theta_{\exp\xi} \circ d(\exp_\xi)\right) = \mathrm{id} \tag{1.12}$$

where τ is as defined above in (1.3). Indeed, we claim that (1.12) is an immediate consequence of (1.11).

Recall the definition (1.3) of the function τ as $\tau(z) = 1/\phi(-z)$. Multiply both sides of (1.11) by $\tau(\operatorname{ad} C(t))$ to obtain

$$\tau(\operatorname{ad} C(t)) \exp(-C(t)) \frac{d}{dt} \exp(C(t)) = C'(t). \tag{1.13}$$

Choose the curve C so that $\xi = C(0)$ and $\eta = C'(0)$. Then the chain rule says that

$$\frac{d}{dt} \exp(C(t))_{|t=0} = d(\exp)_\xi(\eta).$$

Thus

$$\left(\exp(-C(t)) \frac{d}{dt} \exp(C(t))\right)_{|t=0} = \theta_{\exp\xi} d(\exp)_\xi \eta,$$

the result of applying the Maurer-Cartan form θ (at the point $\exp(\xi)$) to the image of η under the differential of exponential map at $\xi \in \mathbf{g}$. Then (1.13) at $t = 0$ translates into (1.12). QED

1.6 The averaging method.

In this section we will give another important application of (1.10): For fixed $\xi \in \mathbf{g}$, the differential of the exponential map is a linear map from $\mathbf{g} = T_\xi(\mathbf{g})$ to $T_{\exp\xi}G$. The (differential of) left translation by $\exp\xi$ carries $T_{\exp\xi}(G)$ back to $T_e G = \mathbf{g}$. Let us denote this composite by $\exp_\xi^{-1} d(\exp)_\xi$. So

$$\theta_{\exp\xi} \circ d(\exp)_\xi = d\exp_\xi^{-1} d(\exp)_\xi : \quad \mathbf{g} \to \mathbf{g}$$

is a linear map. We claim that for any $\eta \in \mathbf{g}$

$$\exp_\xi^{-1} d(\exp)_\xi(\eta) = \int_0^1 \operatorname{Ad}_{\exp(-s\xi)} \eta\, ds. \tag{1.14}$$

We will prove this by applying(1.10) to

$$g(s,t)=\exp\left(t(\xi+s\eta)\right).$$

Indeed,

$$\beta(s,t):=g(s,t)^{-1}\frac{\partial g}{\partial t}=\xi+s\eta$$

so

$$\frac{\partial\beta}{\partial s}\equiv\eta$$

and

$$\beta(0,t)\equiv\xi.$$

The left hand side of (1.14) is $\alpha(0,1)$ where

$$\alpha(s,t):=g(s,t)^{-1}\frac{\partial g}{\partial s}$$

so we may use (1.10) to get an ordinary differential equation for $\alpha(0,t)$. Defining

$$\gamma(t):=\alpha(0,t),$$

(1.10) becomes

$$\frac{d\gamma}{dt}=\eta+[\gamma,\xi]. \tag{1.15}$$

For any $\zeta\in\mathbf{g}$,

$$\begin{aligned}\frac{d}{dt}\mathrm{Ad}_{\exp -t\xi}\zeta &= \mathrm{Ad}_{\exp -t\xi}[\zeta,\xi]\\ &= [\mathrm{Ad}_{\exp -t\xi}\zeta,\xi].\end{aligned}$$

So for constant $\zeta\in\mathbf{g}$,

$$\mathrm{Ad}_{\exp -t\xi}\zeta$$

is a solution of the homogeneous equation corresponding to (1.15). So, by Lagrange's method of variation of constants, we look for a solution of (1.15) of the form

$$\gamma(t)=\mathrm{Ad}_{\exp -t\xi}\zeta(t)$$

and (1.15) becomes

$$\zeta'(t)=\mathrm{Ad}_{\exp t\xi}\eta$$

or

$$\gamma(t)=\mathrm{Ad}_{\exp -t\xi}\int_0^t\mathrm{Ad}_{\exp s\xi}\eta ds$$

is the solution of (1.15) with $\gamma(0)=0$. Setting $s=1$ gives

$$\gamma(1)=\mathrm{Ad}_{\exp -\xi}\int_0^1\mathrm{Ad}_{\exp s\xi}ds$$

and replacing s by $1-s$ in the integral gives (1.14).

1.7 The Euler MacLaurin Formula.

We pause to remind the reader of a different role that the τ function plays in mathematics. We have seen in (1.12) that τ enters into the inverse of the exponential map. In a sense, this formula is taking into account the non-commutativity of the group multiplication, so τ is helping to relate the non-commutative to the commutative.

But much earlier in mathematical history, τ was introduced to relate the discrete to the continuous: Let D denote the differentiation operator in one variable. Then if we think of D as the one dimensional vector field $\partial/\partial h$ it generates the one parameter group $\exp hD$ which consists of translation by h. In particular, taking $h = 1$ we have

$$\left(e^D f\right)(x) = f(x+1).$$

This equation is equally valid in a purely algebraic sense, taking f to be a polynomial and

$$e^D = 1 + D + \frac{1}{2}D^2 + \frac{1}{3!}D^3 + \cdots .$$

This series is infinite. But if p is a polynomial of degree d, then $D^k p = 0$ for $k > D$ so when applied to any polynomial, the above sum is really finite. Since

$$D^k e^{ah} = a^k e^{ah}$$

it follows that if F is any formal power series in one variable, we have

$$F(D)e^{ah} = F(a)e^{ah} \tag{1.16}$$

in the ring of power series in two variables. Of course, under suitable convergence conditions this is an equality of functions of h.

For example, the function $\tau(z) = z/(1 - e^{-z})$ converges for $|z| < 2\pi$ since $\pm 2\pi i$ are the closest zeros of the denominator (other than 0) to the origin. Hence

$$\tau\left(\frac{d}{dh}\right)\frac{e^{zh}}{z} = e^{zh}\frac{1}{1 - e^{-z}} \tag{1.17}$$

holds for $0 < |z| < 2\pi$. Here the infinite order differential operator on the left is regarded as the limit of the finite order differential operators obtained by truncating the power series for τ at higher and higher orders.

Let $a < b$ be integers. Then for any non-negative values of h_1 and h_2 we have

$$\int_{a-h_1}^{b+h_2} e^{zx}dx = e^{h_2 z}\frac{e^{bz}}{z} - e^{-h_2 z}\frac{e^{az}}{z}$$

for $z \neq 0$. So if we set

$$D_1 := \frac{d}{dh_1}, \quad D_2 := \frac{d}{dh_2},$$

the for $0 < |z| < 2\pi$ we have

$$\tau(D_1)\tau(D_2)\int_{a-h_1}^{b+h_2} e^{zx}dx = \tau(z)e^{h_2z}\frac{e^{bz}}{z} - \tau(-z)e^{-h_1z}\frac{e^{az}}{z}$$

because $\tau(D_1)f(h_2) = f(h_2)$ when applied to any function of h_2 since the constant term in τ is one and all of the differentiations with respect to h_1 give zero.
Setting $h_1 = h_2 = 0$ gives

$$\tau(D_1)\tau(D_2)\int_{a-h_1}^{b+h_2} e^{zx}dx\Bigg|_{h_1=h_2=0} = \frac{e^{az}}{1-e^z} + \frac{e^{bz}}{1-e^{-z}}, 0 < |z| < 2\pi.$$

On then other hand, the geometric sum gives

$$\sum_{k=a}^{b} e^{kz} = e^{az}\left(1 + e^z + e^{2z} + \cdots + e^{(b-a)z}\right) = e^{az}\frac{1-e^{(b-a+1)z}}{1-e^z}$$

$$= \frac{e^{az}}{1-e^z} + \frac{e^{bz}}{1-e^{-z}}.$$

We have thus proved the following exact Euler-MacLaurin formula:

$$\tau(D_1)\tau(D_2)\int_{a-h_1}^{b+h_2} f(x)dx\Bigg|_{h_1=h_2=0} = \sum_{k=a}^{b} f(k), \tag{1.18}$$

where the sum on the right is over integer values of k and we have proved this formula for functions f of the form $f(x) = e^{zx}$, $0 < |z| < 2\pi$. It is also true when $z = 0$ by passing to the limit or by direct evaluation.

Repeatedly differentiating (1.18) (with $f(x) = e^{zx}$) with respect to z gives the corresponding formula with $f(x) = x^n e^{zx}$ and hence for all functions of the form $x \mapsto p(x)e^{zx}$ where p is a polynomial and $|z| < 2\pi$.

There is a corresponding formula with remainder for C^k functions.

1.8 The universal enveloping algebra.

We will now give an alternative (algebraic) version of the Campbell-Baker-Hausdorff theorem. It depends on several notions which are extremely important in their own right, so we pause to develop them.

A **universal algebra** of a Lie algebra L is a map $\epsilon : L \to UL$ where UL is an associative algebra with unit such that

1. ϵ is a Lie algebra homomorphism, i.e. it is linear and

$$\epsilon[x, y] = \epsilon(x)\epsilon(y) - \epsilon(y)\epsilon(x)$$

2. If A is any associative algebra with unit and $\alpha : L \to A$ is any Lie algebra homomorphism then there exists a unique homomorphism ϕ of associative algebras such that
$$\alpha = \phi \circ \epsilon.$$

It is clear that if UL exists, it is unique up to a unique isomorphism. So we may then talk of *the* universal algebra of L. We will call it the universal enveloping algebra and sometimes put in parenthesis, i.e. write $U(L)$.

In case $L = \mathbf{g}$ is the Lie algebra of left invariant vector fields on a group G, we may think of L as consisting of left invariant first order homogeneous differential operators on G. Then we may take UL to consist of all left invariant differential operators on G. In this case the construction of UL is intuitive and obvious. The ring of differential operators $\mathcal{D}$ on any manifold is filtered by degree: $\mathcal{D}^n$ consisting of those differential operators with total degree at most n. The quotient, $\mathcal{D}^n/\mathcal{D}^{n-1}$ consists of those homogeneous differential operators of degree n, i.e. homogeneous polynomials in the vector fields with function coefficients. For the case of left invariant differential operators on a group, these vector fields may be taken to be left invariant, and the function coefficients to be constant. In other words, $(UL)^n/(UL)^{n-1}$ consists of all symmetric polynomial expressions, homogeneous of degree n in L. This is the content of the Poincaré-Birkhoff-Witt theorem. In the algebraic case we have to do some work to get all of this. We first must construct $U(L)$.

1.8.1 Tensor product of vector spaces.

Let $E_1, \ldots, E_m$ be vector spaces and (f, F) a multilinear map $f : E_1 \times \cdots \times E_m \to F$. Similarly (g, G). If ℓ is a linear map $\ell : F \to G$, and $g = \ell \circ f$ then we say that ℓ is a morphism of (f, F) to (g, G). In this way we make the set of all (f, F) into a *category.* Want a universal object in this category; that is, an object with a unique morphism into every other object. So want a pair $(t, \mathcal{T})$ where $\mathcal{T}$ is a vector space, $t : E_1 \times \cdots \times E_m \to \mathcal{T}$ is a multilinear map, and for every (f, F) there is a unique linear map $\ell_f : \mathcal{T} \to F$ with
$$f = \ell_f \circ t$$

.

Uniqueness. By the universal property $t = \ell'_t \circ t'$, $t' = \ell'_t \circ t$ so $t = (\ell'_t \circ \ell_{t'}) \circ t$, but also $t = t \circ \mathrm{id}$. So $\ell'_t \circ \ell_{t'} = \mathrm{id}$. Similarly the other way. Thus $(t, \mathcal{T})$, if it exists, is unique up to a unique morphism. This is a standard argument valid in any category proving the uniqueness of "initial elements".

Existence. Let M be the free vector space on the symbols $x_1, \ldots, x_m,\ \ x_i \in E_i$. Let N be the subspace generated by all the
$$(x_1, \ldots, x_i + x'_i, \ldots, x_m) - (x_1, \ldots, x_i, \ldots, x_m) - (x_1, \ldots, x'_i, \ldots, x_m)$$
and all the
$$(x_1, \ldots, , ax_i, \ldots, x_m) - a(x_1, \ldots, x_i, \ldots, x_m)$$

for all $i = 1, \dots, m, x_i, x_i' \in E_i,\ a \in k$. Let $\mathcal{T} = M/N$ and

$$t((x_1, \dots, x_m)) = (x_1, \dots, x_m)/N.$$

This is universal by its very construction. QED

We introduce the notation

$$\mathcal{T} = \mathcal{T}(E_1 \times \cdots \times E_m) =: E_1 \otimes \cdots \otimes E_m.$$

The universality implies an isomorphism

$$(E_1 \otimes \cdots \otimes E_m) \otimes (E_{m+1} \otimes \cdots \otimes E_{m+n}) \cong E_1 \otimes \cdots \otimes E_{m+n}.$$

1.8.2 The tensor product of two algebras.

If A and B are algebras, they are they are vector spaces, so we can form their tensor product as vector spaces. We define a product structure on $A \otimes B$ by defining

$$(a_1 \otimes b_1) \cdot (a_2 \otimes b_2) := a_1 a_2 \otimes b_1 b_2.$$

It is easy to check that this extends to give an algebra structure on $A \otimes B$. In case A and B are associative algebras so is $A \otimes B$, and if in addition both A and B have unit elements, then $1_A \otimes 1_B$ is a unit element for $A \otimes B$. We will frequently drop the subscripts on the unit elements, for it is easy to see from the position relative to the tensor product sign the algebra to which the unit belongs. In other words, we will write the unit for $A \otimes B$ as $1 \otimes 1$. We have an isomorphism of A into $A \otimes B$ given by

$$a \mapsto a \otimes 1$$

when both A and B are associative algebras with units. Similarly for B. Notice that

$$(a \otimes 1) \cdot (1 \otimes b) = a \otimes b = (1 \otimes b) \cdot (a \otimes 1).$$

In particular, an element of the form $a \otimes 1$ commutes with an element of the form $1 \otimes b$.

1.8.3 The tensor algebra of a vector space.

Let V be a vector space. The **tensor algebra** of a vector space is the solution of the universal problem for maps α of V into an associative algebra: it consists of an algebra TV and a map $\iota : V \to TV$ such that ι is linear, and for any linear map $\alpha : V \to A$ where A is an associative algebra there exists a unique algebra homomorphism $\psi : TV \to A$ such that $\alpha = \psi \circ \iota$. We set

$$T^n V := V \otimes \cdots \otimes V \quad n-\text{factors}.$$

We define the multiplication to be the isomorphism

$$T^n V \otimes T^m V \to T^{n+m} V$$

obtained by "dropping the parentheses," i.e. the isomorphism given at the end of the last subsection. Then

$$TV := \bigoplus T^n V$$

(with $T^0 V$ the ground field) is a solution to this universal problem, and hence the unique solution.

1.8.4 Construction of the universal enveloping algebra.

If we take $V = L$ to be a Lie algebra, and let I be the two sided ideal in TL generated the elements $[x, y] - x \otimes y + y \otimes x$ then

$$UL := TL/I$$

is a universal algebra for L. Indeed, any homomorphism α of L into an associative algebra A extends to a unique algebra homomorphism $\psi : TL \to A$ which must vanish on I if it is to be a Lie algebra homomorphism.

1.8.5 Extension of a Lie algebra homomorphism to its universal enveloping algebra.

If $h : L \to M$ is a Lie algebra homomorphism, then the composition

$$\epsilon_M \circ h : L \to UM$$

induces a homomorphism

$$UL \to UM$$

and this assignment sending Lie algebra homomorphisms into associative algebra homomorphisms is functorial.

1.8.6 Universal enveloping algebra of a direct sum.

Suppose that: $L = L_1 \oplus L_2$, with $\epsilon_i : L_i \to U(L_i)$, and $\epsilon : L \to U(L)$ the canonical homomorphisms. Define

$$f : L \to U(L_1) \otimes U(L_2), \quad f(x_1 + x_2) = \epsilon_1(x_1) \otimes 1 + 1 \otimes \epsilon_2(x_2).$$

This is a homomorphism because x_1 and x_2 commute. It thus extends to a homomorphism

$$\psi : U(L) \to U(L_1) \otimes U(L_2).$$

Also,

$$x_1 \mapsto \epsilon(x_1)$$

is a Lie algebra homomorphism of $L_1 \to U(L)$ which thus extends to a unique algebra homomorphism

$$\phi_1 : U(L_1) \to U(L)$$

and similarly $\phi_2 : U(L_2) \to U(L)$. We have

$$\phi_1(x_1)\phi_2(x_2) = \phi_2(x_2)\phi_1(x_1), \quad x_1 \in L_1, x_2 \in L_2$$

since $[x_1, x_2] = 0$. As the $\epsilon_i(x_i)$ generate $U(L_i)$, the above equation holds with x_i replaced by arbitrary elements $u_i \in U(L_i), i = 1, 2$. So we have a homomorphism

$$\phi : U(L_1) \otimes U(L_2) \to U(L), \quad \phi(u_1 \otimes u_2) := \phi_1(u_1)\phi_2(u_2).$$

We have

$$\phi \circ \psi(x_1 + x_2) = \phi(x_1 \otimes 1) + \phi(1 \otimes x_2) = x_1 + x_2$$

so $\phi \circ \psi = \mathrm{id}$, on L and hence on $U(L)$ and

$$\psi \circ \phi(x_1 \otimes 1 + 1 \otimes x_2) = x_1 \otimes 1 + 1 \otimes x_2$$

so $\psi \circ \phi = \mathrm{id}$ on $L_1 \otimes 1 + 1 \otimes L_2$ and hence on $U(L_1) \otimes U(L_2)$. Thus

$$U(L_1 \oplus L_2) \cong U(L_1) \otimes U(L_2).$$

1.8.7 Bialgebra structure.

Consider the map $L \to U(L) \otimes U(L)$:

$$x \mapsto x \otimes 1 + 1 \otimes x.$$

Then

$$(x \otimes 1 + 1 \otimes x)(y \otimes 1 + 1 \otimes y) =$$

$$xy \otimes 1 + x \otimes y + y \otimes x + +1 \otimes xy,$$

and multiplying in the reverse order and subtracting gives

$$[x \otimes 1 + 1 \otimes x, y \otimes 1 + 1 \otimes y] = [x, y] \otimes 1 + 1 \otimes [x, y].$$

Thus the map $x \mapsto x \otimes 1 + 1 \otimes x$ determines an algebra homomorphism

$$\Delta : U(L) \to U(L) \otimes U(L).$$

Define

$$\varepsilon : U(L) \to k, \ \ \varepsilon(1) = 1, \ \ \varepsilon(x) = 0, x \in L$$

and extend as an algebra homomorphism. Then

$$(\varepsilon \otimes \mathrm{id})(x \otimes 1 + 1 \otimes x) = 1 \otimes x, \quad x \in L.$$

We identify $k \otimes L$ with L and so can write the above equation as

$$(\varepsilon \otimes \mathrm{id})(x \otimes 1 + 1 \otimes x) = x, \quad x \in L.$$

The algebra homomorphism

$$(\varepsilon \otimes \mathrm{id}) \circ \Delta : U(L) \to U(L)$$

is the identity (on 1 and on) L and hence is the identity. Similarly

$$(\mathrm{id} \otimes \varepsilon) \circ \Delta = \mathrm{id}.$$

A vector space C with a map $\Delta : C \to C \otimes C$, (called a **comultiplication**) and a map $\varepsilon : D \to k$ (called a **co-unit**) satisfying

$$(\varepsilon \otimes \mathrm{id}) \circ \Delta = \mathrm{id}$$

and

$$(\mathrm{id} \otimes \varepsilon) \circ \Delta = \mathrm{id}$$

is called a **co-algebra**. If C is an algebra and both Δ and ε are algebra homomorphisms, we say that C is a **bi-algebra** (sometimes shortened to "bigebra"). So we have proved that $(U(L), \Delta, \varepsilon)$ is a bialgebra.

Also

$$[(\Delta \otimes \mathrm{id}) \circ \Delta](x) = x \otimes 1 \otimes 1 + 1 \otimes x \otimes 1 + 1 \otimes 1 \otimes x = [(\mathrm{id} \otimes \Delta) \circ \ \Delta](x)$$

for $x \in L$ and hence for all elements of $U(L)$. Hence the comultiplication is is coassociative. (It is also co-commutative.)

1.9 The Poincaré-Birkhoff-Witt Theorem.

Suppose that V is a vector space made into a Lie algebra by declaring that all brackets are zero. Then the ideal I in TV defining $U(V)$ is generated by $x \otimes y - y \otimes x$, and the quotient TV/I is just the symmetric algebra, SV. So the universal enveloping algebra of the trivial Lie algebra is the symmetric algebra.

For any Lie algebra L define U_nL to be the subspace of UL generated by products of at most n elements of L, i.e. by all products

$$\epsilon(x_1) \cdots \epsilon(x_m), \qquad m \leq n.$$

For example,,

$$U_0L = k, \text{ the ground field}$$

and

$$U_1L = k \oplus \epsilon(L).$$

We have

$$U_0L \subset U_1L \subset \cdots \subset U_nL \subset U_{n+1}L \subset \cdots$$

and

$$U_mL \cdot U_mL \subset U_{m+n}L.$$

We define

$$\operatorname{gr}_n UL := U_nL/U_{n-1}L$$

and

$$\operatorname{gr} UL := \bigoplus \operatorname{gr}_n UL$$

with the multiplication

$$\operatorname{gr}_m UL \times \operatorname{gr}_n UL \to \operatorname{gr}_{m+n} UL$$

induced by the multiplication on UL.

If $a \in U_nL$ we let $\overline{a} \in \operatorname{gr}_n UL$ denote its image by the projection $U_nL \to U_nL/U_{n-1}L = \operatorname{gr}_n UL$. We may write a as a sum of products of at most n elements of L:

$$a = \sum_{m_\mu \leq n} c_\mu \epsilon(x_{\mu,1}) \cdots \epsilon(x_{\mu,m_\mu}).$$

Then $\overline{a}$ can be written as the corresponding homogeneous sum

$$\overline{a} = \sum_{m_\mu = n} c_\mu \overline{\epsilon(x_{\mu,1})} \cdots \overline{\epsilon(x_{\mu,m_\mu})}.$$

In other words, as an algebra, $\operatorname{gr} UL$ is generated by the elements $\overline{\epsilon(x)}, \quad x \in L$. But all such elements commute. Indeed, for $x, y \in L$,

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

by the defining property of the universal enveloping algebra. The right hand side of this equation belongs to U_1L. Hence

$$\overline{\epsilon(x)\epsilon(y)} - \overline{\epsilon(y)\epsilon(x)} = 0$$

in $\operatorname{gr}_2 UL$. This proves that $\operatorname{gr} UL$ is *commutative*. Hence, by the universal property of the symmetric algebra, there exists a unique algebra homomorphism

$$\mathbf{w} : SL \to UL$$

extending the linear map

$$L \to \operatorname{gr} UL, \quad x \mapsto \overline{\epsilon(x)}.$$

Since the $\overline{\epsilon(x)}$ generate $\operatorname{gr} UL$ as an algebra, we know that this map is surjective. The **Poincaré-Birkhoff-Witt theorem** asserts that

$$\mathbf{w} : SL \to \operatorname{gr} UL \ \text{ is an isomorphism.} \tag{1.19}$$

Suppose that we choose a basis $x_i, \ i \in I$ of L where I is a totally ordered set. Since

$$\overline{\epsilon(x_i)\epsilon(x_j)} = \overline{\epsilon(x_j)\epsilon(x_i)}$$

we can rearrange any product of $\overline{\epsilon(x_i)}$ so as to be in increasing order. This shows that the elements

$$x_M := \epsilon(x_{i_1}) \cdots \epsilon(x_{i_m}), \quad M := (i_1, \ldots, i_m) \ i_1 \leq \cdots i_m$$

span UL as a vector space. We claim that (1.19) is equivalent to

Theorem 1 Poincaré-Birkhoff-Witt. *The elements x_M form a basis of UL.*

Proof that (1.19) is equivalent to the statement of the theorem. For any expression x_M as above, we denote its length by $\ell(M) = m$. The elements $\overline{x_M}$ are the images under $\mathbf{w}$ of the monomial basis in $S_m(L)$. As we know that $\mathbf{w}$ is surjective, equation (1.19) is equivalent to the assertion that $\mathbf{w}$ is injective. This amounts to the non-existence of a relation of the form

$$\sum_{\ell(M)=n} c_M x_M = \sum_{\ell(M)<n} c_M x_M$$

with some non-zero coefficients on the left hand side. But any non-trivial relation between the x_M can be rewritten in the above form by moving the terms of highest length to one side. QED

We now turn to the proof of the theorem:

Let V be the vector space with basis z_M where M runs over all ordered sequences $i_1 \leq i_2 \leq \cdots \leq i_n$. (Recall that we have chosen a well ordering on I and that the $x_i\ {}_{i\in I}$ form a basis of L.)

Furthermore, the empty sequence, $z_\emptyset$ is allowed, and we will identify the symbol $z_\emptyset$ with the number $1 \in k$. If $i \in I$ and $M = (i_1, \ldots, i_n)$ we write $i \leq M$ if $i \leq i_1$ and then let (i, M) denote the ordered sequence $(i, i_1, \ldots, i_n)$. In particular, we adopt the convention that if $M = \emptyset$ is the empty sequence then $i \leq M$ for all i in which case $(i, M) = (i)$. Recall that if $M = (i_1, \ldots, i_n))$ we set $\ell(M) = n$ and call it the length of M. So, for example, $\ell(i, M) = \ell(M) + 1$ if $i \leq M$.

Lemma 1 *We can make V into an L module in such a way that*

$$x_i z_M = z_{iM} \quad \textit{whenever } i \leq M. \tag{1.20}$$

Proof of lemma. We will inductively define a map

$$L \times V \to V, \quad (x, v) \mapsto xv$$

and then show that it satisfies the equation

$$xyv - yxv = [x, y]v, \quad x, y \in g, \quad v \in V, \tag{1.21}$$

which is the condition that makes V into an L module. Our definition will be such that (1.20) holds. In fact, we will define $x_i z_M$ inductively on $\ell(M)$ and on i. So we start by defining

$$x_i z_\emptyset = z_{(i)}$$

which is in accordance with (1.20). This defines $x_i z_M$ for $\ell(M) = 0$. For $\ell(M) = 1$ we define

$$x_i z_{(j)} = z_{(i,j)} \quad \text{if } i \leq j$$

while if $i > j$ we set

$$x_i z_{(j)} = x_j z_{(i)} + [x_i, x_j] z_\emptyset = z_{(j,i)} + \sum c_{ij}^k z_{(k)}$$

where

$$[x_i, x_j] = \sum c_{ij}^k x_k$$

is the expression for the Lie bracket of x_i with x_j in terms of our basis. These c_{ij}^k are known as the **structure constants** of the Lie algebra, L in terms of the given basis. Notice that the first of these two cases is consistent with (and forced on us) by (1.20) while the second is forced on us by (1.21). We now have defined $x_i z_M$ for all i and all M with $\ell(M) \leq 1$. and we have done so in such a way that (1.20) holds, and (1.21) holds where it makes sense (i.e. for $\ell(M) = 0$).

So suppose that we have defined $x_j z_N$ for all j if $\ell(N) < \ell(M)$ and for all $j < i$ if $\ell(N) = \ell(M)$ in such a way that

$$x_j z_N \text{ is a linear combination of } z_L\text{'s with } \ell(L) \leq \ell(N) + 1 \quad (*).$$

We then define

$$x_i z_M = \begin{cases} z_{iM} \text{ if } i \leq M \\ x_j(x_i z_N) + [x_i, x_j] z_N \text{ if } M = (jN) \text{ with } i > j. \end{cases} \tag{1.22}$$

This makes sense since $x_i z_N$ is already defined as a linear combination of z_L's with $\ell(L) \leq \ell(N) + 1 = \ell(M)$ and because $[x_i, x_j]$ can be written as a linear combination of the x_k as above. Furthermore $(*)$ holds with j and N replaced by M. Furthermore, (1.20) holds by construction. We must check (1.21). By linearity, this means that we must show that

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

If $i = j$ both sides are zero. Also, since both sides are anti-symmetric in i and j, we may assume that $i > j$. If $j \leq N$ and $i > j$ then this equation holds by definition. So we need only deal with the case where $j \not\leq N$ which means that $N = (kP)$ with $k \leq P$ and $i > j > k$. So we have, by definition,

$$\begin{aligned} x_j z_N &= x_j z_{(kP)} \\ &= x_j x_k z_P \\ &= x_k x_j z_P + [x_j, x_k] z_P. \end{aligned}$$

Now if $j \leq P$ then $x_j z_P = z_{(jP)}$ and $k < (jP)$. If $j \not\leq P$ then $x_j z_P = z_Q + w$ where still $k \leq Q$ and w is a linear combination of elements of length $< \ell(N)$. So we know that (1.21) holds for $x = x_i, y = x_k$ and $v = z_{(jP)}$ (if $j \leq P$) or $v = z_Q$ (otherwise). Also, by induction, we may assume that we have verified (1.21) for all N' of length $< \ell(N)$. So we may apply (1.21) to $x = x_i, \quad y = x_k$ and $v = x_j z_P$ and also to $x = x_i, \ y = [x_j, x_k], \ v = z_P$. So

$$x_i x_j z_N = x_k x_i x_j z_P + [x_i, x_k] x_j z_P + [x_j, x_k] x_i z_P + [x_i, [x_j, x_k]] z_P.$$

Similarly, the same result holds with i and j interchanged. Subtracting this interchanged version from the preceding equation the two middle terms from each equation cancel and we get

$$\begin{aligned}(x_ix_j - x_jx_i)z_N &= x_k(x_ix_j - x_jx_i)z_P + ([x_i,[x_j,x_k]] - [x_j,[x_i,x_k]])z_P\\ &= x_k[x_i,x_j]z_P + ([x_i,[x_j,x_k]] - [x_j,[x_i,x_k]])z_P\\ &= [x_i,x_j]x_kz_P + ([x_k,[x_i,x_j]] + [x_i,[x_j,x_k]] - [x_j,[x_i,x_k]])z_P\\ &= [x_i,x_j]z_N.\end{aligned}$$

(In passing from the second line to the third we used (1.21) applied to z_P (by induction) and from the third to the last we used the antisymmetry of the bracket and Jacobi's equation.)QED

Proof of the PBW theorem. We have made V into an L and hence into a $U(L)$ module. By construction, we have, inductively,

$$x_Mz_\emptyset = z_M.$$

But if

$$\sum c_Mx_M = 0$$

then

$$0 = \sum c_Mz_M = \left(\sum c_Mx_M\right)z_\emptyset$$

contradicting the fact the the z_M are independent. QED

In particular, the map $\epsilon : L \to U(L)$ is an injection, and so we may identify L as a subspace of $U(L)$.

1.10 Primitives.

An element x of a bialgebra is called **primitive** if

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

So the elements of L are primitives in $U(L)$.

We claim that *these are the only primitives.*

First prove this for the case L is abelian so $U(L) = S(L)$. Then we may think of $S(L) \otimes S(L)$ as polynomials in twice the number of variables as those of $S(L)$ and

$$\Delta(f)(u,v) = f(u+v).$$

The condition of being primitive says that

$$f(u+v) = f(u) + f(v).$$

Taking homogeneous components, the same equality holds for each homogeneous component. But if f is homogeneous of degree n, taking $u = v$ gives

$$2^nf(u) = 2f(u)$$

so $f = 0$ unless $n = 1$.

Taking gr, this shows that for any Lie algebra the primitives are contained in $U_1(L)$. But

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

so the condition on primitivity requires $c = 2c$ or $c = 0$. QED

1.11 Free Lie algebras

1.11.1 Magmas and free magmas on a set

A set M with a map:

$$M \times M \to M, \quad (x, y) \mapsto xy$$

is called a **magma**. Thus a magma is a set with a binary operation with no axioms at all imposed.

Let X be any set. Define X_n inductively by $X_1 := X$ and

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

for $n \geq 2$. Thus X_2 consists of all expressions ab where a and b are elements of X. (We write ab instead of (a, b).) An element of X_3 is either an expression of the form $(ab)c$ or an expression of the form $a(bc)$. An element of X_4 has one out of five forms: $a((bc)d), a(b(cd)), ((ab)(cd)), ((ab)c)d$ or $(a(bc))d$.

Set

$$M_X := \coprod_{n=1}^{\infty} X_n.$$

An element $w \in M_X$ is called a non-associative word, and its length $\ell(w)$ is the unique n such that $w \in X_n$. We have a "multiplication" map $M_X \times M_X$ given by the inclusion

$$X_p \times X_q \hookrightarrow X_{p+q}.$$

Thus the multiplication on M_X is concatenation of non-associative words.

If N is any magma, and $f : X \to N$ is any map, we define $F : M_X \to N$ by $F = f$ on X_1, by

$$F : X_2 \to N, \quad F(ab) = f(a)f(b)$$

and inductively

$$F : X_p \times X_q \to N, \quad F(uv) = F(u)F(v).$$

Any element of X_n has a unique expression as uv where $u \in X_p$ and $v \in X_q$ for a unique (p, q) with $p + q = n$, so this inductive definition is valid.

It is clear that F is a magna homomorphism and is uniquely determined by the original map f. Thus M_X is the "free magma on X" or the "universal magma on X" in the sense that it is the solution to the universal problem associated to a map from X to any magma.

Let A_X be the vector space of finite formal linear combinations of elements of M_X. So an element of A_X is a finite sum $\sum c_m m$ with $m \in M_X$ and c_m in the ground field. The multiplication in M_X extends by bi-linearity to make A_X into an algebra. If we are given a map $X \to B$ where B is any algebra, we get a unique magna homomorphism $M_X \to B$ extending this map (where we think of B as a magma) and then a unique algebra map $A_X \to B$ extending this map by linearity.

Notice that the algebra A_X is graded since every element of M_X has a length and the multiplication on M_X is graded. Hence A_X is the free algebra on X in the sense that it solves the universal problem associated with maps of X to algebras.

1.11.2 The Free Lie Algebra L_X.

In A_X let I be the two-sided ideal generated by all elements of the form $aa,\ a \in A_X$ and $(ab)c + (bc)a + (ca)b, \quad a, b, c \in A_X$. We set

$$L_X := A_X/I$$

and call L_X the free Lie algebra on X. Any map from X to a Lie algebra L extends to a unique algebra homomorphism from L_X to L.

We claim that the ideal I defining L_X is graded. This means that if $a = \sum a_n$ is a decomposition of an element of I into its homogeneous components, then each of the a_n also belong to I. To prove this, let $J \subset I$ denote the set of all $a = \sum a_n$ with the property that all the homogeneous components a_n belong to I. Clearly J is a two sided ideal. We must show that $I \subset J$. For this it is enough to prove the corresponding fact for the generating elements. Clearly if

$$a = \sum a_p,\ \ b = \sum b_q, c = \sum c_r$$

then

$$(ab)c + (bc)a + (ca)b = \sum_{p,q,r} \left((a_p b_q)c_r + (b_q c_r)a_p + (c_r a_p)b_q\right).$$

But also if $x = \sum x_m$ then

$$x^2 = \sum x_n^2 + \sum_{m<n} (x_m x_n + x_n x_m)$$

and

$$x_m x_n + x_n x_m = (x_m + x_n)^2 - x_m^2 - x_n^2 \in I$$

so $I \subset J$.

The fact that I is graded means that L_X inherits the structure of a graded algebra.

1.11.3 The free associative algebra Ass(X).

Let V_X be the vector space of all finite formal linear combinations of elements of X. Define

$$\mathrm{Ass}_X = T(V_X),$$

the tensor algebra of V_X. Any map of X into an associative algebra A extends to a unique linear map from V_X to A and hence to a unique algebra homomorphism from Ass_X to A. So Ass_X is the free associative algebra on X.

We have the maps $X \to L_X$ and $\epsilon : L_X \to U(L_X)$ and hence their composition maps X to the associative algebra $U(L_X)$ and so extends to a unique homomorphism

$$\Psi : \mathrm{Ass}_X \to U(L_X).$$

On the other hand, the commutator bracket gives a Lie algebra structure to Ass_X and the map $X \to \mathrm{Ass}_X$ thus give rise to a Lie algebra homomorphism

$$L_X \to \mathrm{Ass}_X$$

which determines an associative algebra homomorphism

$$\Phi : U(L_X) \to \mathrm{Ass}_X.$$

both compositions $\Phi \circ \Psi$ and $\Psi \circ \Phi$ are the identity on X and hence, by uniqueness, the identity everywhere. We obtain the important result that $U(L_X)$ and Ass_X are canonically isomorphic:

$$U(L_X) \cong \mathrm{Ass}_X. \tag{1.23}$$

Now the Poincaré-Birkhoff -Witt theorem guarantees that the map $\epsilon : L_X \to U(L_X)$ is injective. So under the above isomorphism, the map $L_X \to \mathrm{Ass}_X$ is injective. On the other hand, by construction, the map $X \to V_X$ induces a surjective Lie algebra homomorphism from L_X into the Lie subalgebra of Ass_X generated by X. So we see that the under the isomorphism (1.23) $L_X \subset U(L_X)$ is mapped isomorphically onto the Lie subalgebra of Ass_X generated by X.

Now the map

$$X \to \mathrm{Ass}_X \otimes \mathrm{Ass}_X, \quad x \mapsto x \otimes 1 + 1 \otimes x$$

extends to a unique algebra homomorphism

$$\Delta : \mathrm{Ass}_X \to \mathrm{Ass}_X \otimes \mathrm{Ass}_X.$$

Under the identification (1.23) this is none other than the map

$$\Delta : U(L_X) \to U(L_X) \otimes U(L_X)$$

and hence we conclude that L_X is the set of primitive elements of Ass_X:

$$L_X = \{w \in \mathrm{Ass}_X | \Delta(w) = w \otimes 1 + 1 \otimes w.\} \tag{1.24}$$

under the identification (1.23).

1.12 Algebraic proof of CBH and explicit formulas.

We recall our constructs of the past few sections: X denotes a set, L_X the free Lie algebra on X and Ass_X the free associative algebra on X so that Ass_X may be identified with the universal enveloping algebra of L_X. Since Ass_X may be identified with the non-commutative polynomials indexed by X, we may consider its completion, F_X, the algebra of formal power series indexed by X. Since the free Lie algebra L_X is graded we may also consider its completion which we shall denote by $\mathbf{L}_X$. Finally let m denote the ideal in F_X generated by X. The maps

$$\exp : m \to 1+m, \quad \log : 1+m \to m$$

are well defined by their formal power series and are mutual inverses. (There is no convergence issue since everything is within the realm of formal power series.) Furthermore exp is a bijection of the set of $\alpha \in m$ satisfying $\Delta\alpha = \alpha\otimes 1 + 1\otimes\alpha$ to the set of all $\beta \in 1+m$ satisfying $\Delta\beta = \beta\otimes\beta$.

1.12.1 Abstract version of CBH and its algebraic proof.

In particular, since the set $\{\beta \in 1+m | \Delta\beta = \beta\otimes\beta\}$ forms a group, we conclude that for any $A, B \in \mathbf{L}_X$ there exists a $C \in \mathbf{L}_X$ such that

$$\exp C = (\exp A)(\exp B).$$

This is the abstract version of the Campbell-Baker-Hausdorff formula. It depends basically on two algebraic facts: That the universal enveloping algebra of the free Lie algebra is the free associative algebra, and that the set of primitive elements in the universal enveloping algebra (those satisfying $\Delta\alpha = \alpha\otimes 1 + 1\otimes\alpha$) is precisely the original Lie algebra.

1.12.2 Explicit formula for CBH.

Define the map

$$\Phi : m \cap \mathrm{Ass}_X \to L_X,$$

$$\Phi(x_1 \dots x_n) := [x_1, [x_2, \dots, [x_{n-1}, x_n] \cdots] = \mathrm{ad}(x_1) \cdots \mathrm{ad}(x_{n-1})(x_n),$$

and let $\Theta : \mathrm{Ass}_X \to \mathrm{End}(L_X)$ be the algebra homomorphism extending the Lie algebra homomorphism $\mathrm{ad} : L_X \to \mathrm{End}(L_X)$. We claim that

$$\Phi(uv) = \Theta(u)\Phi(v), \quad \forall\ u \in \mathrm{Ass}_X,\ v \in m \cap \mathrm{Ass}_X. \tag{1.25}$$

Proof. It is enough to prove this formula when u is a monomial, $u = x_1 \cdots x_n$. We do this by induction on n. For $n = 0$ the assertion is obvious and for $n = 1$

it follows from the definition of Φ. Suppose $n > 1$. Then

$$\begin{aligned}
\Phi(x_1 \cdots x_n v) &= \Theta(x_1)\Phi(x_2 \cdots x_n v) \\
&= \Theta(x_1)\Theta(x_2 \ldots x_n)\Phi(v) \\
&= \Theta(x_1 \cdots x_n)\Phi(v). \text{ QED}
\end{aligned}$$

Let L_X^n denote the $n-$th graded component of L_X. So L_X^1 consists of linear combinations of elements of X, L_X^2 is spanned by all brackets of pairs of elements of X, and in generalL_X^n is spanned by elements of the form

$$[u, v],\ u \in L_X^p,\ v \in L_X^q,\ p + q = n.$$

We claim that

$$\Phi(u) = nu \ \ \forall\ u \in L_X^n. \tag{1.26}$$

For $n = 1$ this is immediate from the definition of Φ. So by induction it is enough to verify this on elements of the form $[u, v]$ as above. We have

$$\begin{aligned}
\Phi([u, v]) &= \Phi(uv - vu) \\
&= \Theta(u)\Phi(v) - \Theta(v)\Phi(u) \\
&= q\Theta(u)v - p\Theta(v)u \ \text{ by induction} \\
&= q[u, v] - p[v, u] \\
&\qquad \text{since } \Theta(w) = \mathrm{ad}(w) \text{ for } w \in L_X \\
&= (p + q)[u, v] \quad \text{QED.}
\end{aligned}$$

We can now write down an explicit formula for the $n-$th term in the Campbell-Baker-Hausdorff expansion. Consider the case where X consists of two elements $X = \{x, y\},\ x \neq y$. Let us write

$$z = \log\left((\exp x)(\exp y)\right) \quad z \in \mathbf{L}_X, \quad z = \sum_1^\infty z_n(x, y).$$

We want an explicit expression for $z_n(x, y)$. We know that

$$z_n = \frac{1}{n}\Phi(z_n)$$

and z_n is a sum of non-commutative monomials of degree n in x and y. Now

$$\begin{aligned}
(\exp x)(\exp y) &= \left(\sum_{p=0}^{\infty}\frac{x^p}{p!}\right)\left(\sum_{q=0}^{\infty}\frac{y^q}{q!}\right)\\
&= 1+\sum_{p+q\geq 1}\frac{x^p y^q}{p!q!} \quad \text{so}\\
z &= \log((\exp x)(\exp y))\\
&= \sum_{m=1}^{\infty}\frac{(-1)^{m+1}}{m}\left(\sum_{p+q\geq 1}\frac{x^p y^q}{p!q!}\right)^m\\
&= \sum_{p_i+q_i\geq 1}\frac{(-1)^{m+1}}{m}\frac{x^{p_1}y^{q_1}x^{p_2}y^{q_2}\cdots x^{p_m}y^{q_m}}{p_1!q_1!\cdots p_m!q_m!}.
\end{aligned}$$

We want to apply $\frac{1}{n}\Phi$ to the terms in this last expression which are of total degree n so as to obtain z_n. So let us examine what happens when we apply Φ to an expression occurring in the numerator: If $q_m \geq 2$ we get 0 since we will have $\mathrm{ad}(y)(y) = 0$. Similarly we will get 0 if $q_m = 0, p_m \geq 2$. Hence the only terms which survive are those with $q_m = 1$ or $q_m = 0, p_m = 1$. Accordingly we decompose z_n into these two types:

$$z_n = \frac{1}{n}\sum_{p+q=n}(z'_{p,q}+z''_{p,q}), \tag{1.27}$$

where

$$z'_{p,q} = \sum \frac{(-1)^{m+1}}{m}\frac{\mathrm{ad}(x)^{p_1}\mathrm{ad}(y)^{q_1}\cdots\mathrm{ad}(x)^{p_m}y}{p_1!q_1!\cdots p_m!} \quad \text{summed over all}$$

$$p_1+\cdots+p_m = p,\ q_1+\cdots+q_{m-1} = q-1,\quad q_i+p_i\geq 1,\ p_m\geq 1$$

and

$$z''_{p,q} = \sum \frac{(-1)^{m+1}}{m}\frac{\mathrm{ad}(x)^{p_1}\mathrm{ad}(y)^{q_1}\cdots\mathrm{ad}(y)^{q_{m-1}}(x)}{p_1!q_1!\cdots q_{m-1}!} \quad \text{summed over}$$

$$\begin{gathered}
p_1+\cdots+p_{m-1} = p-1,\quad q_1+\cdots+q_{m-1} = q,\\
p_i+q_i\geq 1\ (i=1,\ldots,m-1)\ q_{m-1}\geq 1.
\end{gathered}$$

The first four terms are:

$$\begin{aligned}
z_1(x,y) &= x+y\\
z_2(x,y) &= \frac{1}{2}[x,y]\\
z_3(x,y) &= \frac{1}{12}[x,[x,y]]+\frac{1}{12}[y,[y,x]]\\
z_4(x,y) &= \frac{1}{24}[x,[y,[x,y]]].
\end{aligned}$$