7. Baker-Campbell-Hausdorff formula

7.1. Formulation. Let $G \subset \mathrm{GL}(n, \mathbf{R})$ be a matrix Lie group and let $\mathfrak{g} = \mathrm{Lie}(G)$. The exponential map is an analytic diffeomorphim of a neighborhood of 0 in $\mathfrak{g}$ with a neighborhood of 1 in G. So for $X, Y \in \mathfrak{g}$ sufficiently close to 0 we can write

$$\exp X \exp Y = \exp Z$$

where

$$Z : (X, Y) \longmapsto Z(X, Y) \qquad (|X|, |Y| < a)$$

is an analytic map into $\mathfrak{g}$. It turns out that one can compute this map explicitly. The resulting formula is called the *Baker-Campbell-Hausdorff formula.* For fixed X, Y we replace X and Y by tX and tY so that

$$\exp tX \exp tY = \exp Z(tX, tY) = \exp Z(t, X, Y).$$

Then $Z(t, X, Y)$ is analytic for small t and so we can write

$$Z(t, X, Y) = \sum_{n=0}^{\infty} t^n z_n(X, Y)$$

where the series converges for small t. It is almost obvious that z_n is a homogeneous polynomial map of $\mathfrak{g} \times \mathfrak{g}$ into $\mathfrak{g}$ of degree n. But it is far less obvious that it is a *Lie polynomial*, namely, that it is made up of commutators involving X, Y of degree n. This is the essence of the BCH formula. The z_n can be explicitly determined and are given by a famous formula due to Dynkin.

One can have an idea of what is involved by computing z_n for very small n. This is a formal exercise and so we can operate in the associative algebra $\mathcal{P}$ of formal power series in two *non-commutative* indeterminates x, y. Then

$$z = \log(e^x e^y) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n} \left(e^x e^y - 1\right)^n .$$

It is now a question of replacing the exponentials by their power series and calculating the term z_n as the part of the expansion of degree n. It is obvious that z_n is an *associative* polynomial. A straightforward calculation gives (with $[a,b]=ab-ba$ in $\mathcal{P}$)

$$\begin{aligned}
z_0 &= 0\\
z_1 &= x+y\\
z_2 &= \frac{1}{2}[x,y]\\
z_3 &= \frac{1}{12}[x,[x,y]]-\frac{1}{12}[y,[x,y]]\\
z_4 &= -\frac{1}{48}[x,x,[x,y]]]-\frac{1}{48}[y,x,[x,y]]]\text{etc.}
\end{aligned}$$

One can proceed further but the calculations become prohibitive very rapidly unless one uses a computer. That all the z_n are Lie polynomials is thus a formal fact of considerable interest and difficulty to establish.

Our method of attack is to start with the equation

$$\exp tX \exp tY = \exp Z(tX,tY) = \exp Z(t,X,Y).$$

and obtain a differential equation (ordinary, but non-linear) for $Z(t,X,Y)$ from which the z_n can be determined by resursion. The structure of the recursion will make it clear that the z_n are Lie polynomials and that the series

$$Z(X,Y)=\sum_{n=0}^{\infty} z_n(X,Y)$$

will converge for small X,Y. But in order to carry this out we need to know how to compute the derivative of the exponential function e^X at an arbitrary point X. We have done this only at $X=0$ and the result for arbitrary X is much more involved.

7.2. Derivative of e^X at an arbitrary X. Since X is a matrix variable, the derivative has to be along an arbitrary direction Y. We thus wish to determine

$$\left(\frac{d}{dt}\right)_{t=0} e^{X+tY}.$$

We shall transfer the calculation to I by left translating by e^X so that we really calculate

$$D_Y(e^X) := e^{-X}\left(\frac{d}{dt}\right)_{t=0} e^{X+tY}$$

instead of

$$d_Y(e^X) := \left(\frac{d}{dt}\right)_{t=0} e^{X+tY}.$$

The result is the following. Let g be the entire function

$$g(z) = \frac{1-e^{-z}}{z} = \sum_{m=0}^{\infty} \frac{(-1)^m}{(m+1)!} z^m.$$

Theorem 1. *We have*

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

Equivalently

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

Proof. We have

$$\begin{aligned} D_Y(e^X) &= e^{-X} \sum_{n\geq 1} \left(\frac{d}{dt}\right)_{t=0} \frac{(X+tY)^n}{n!} \\ &= \sum_{r\geq 0, n\geq 0, 0\leq s\leq n} \frac{(-1)^r}{r!(n+1)!} X^r X^s Y X^{n-s}. \end{aligned}$$

We go to summation variables m, r, s where $m = r + n$; the ranges are $m \geq 0, r \geq 0, s \geq 0, r + s \leq m$. Thus

$$\begin{aligned} D_Y(e^X) &= \sum_{m,r,s} \frac{(-1)^r}{r!(m+1-r)!} X^{r+s} Y X^{m-r-s} \\ &= \sum_{m\geq 0, 0\leq k\leq m, 0\leq r\leq k} \frac{(-1)^m}{(m+1)!} (-1)^{m-k} \binom{m+1}{r} X^k Y X^{m-k}. \end{aligned}$$

We now use the identity

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

which is easily established by induction on r. Moreover, if λ, ρ are the endomorphisms of the algebra of $n \times n$ matrices of left and right multiplication by X, they commute with each other, $ad\ X = \lambda - \rho$, and

$$
\begin{aligned}
(ad\ X)^m(Y) &= \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} \lambda^k \rho^{m-k}(Y) \\
&= \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} X^k Y X^{m-k}.
\end{aligned}
$$

Hence

$$
D_Y(e^X) = \sum_{m=0}^{\infty} \frac{(-1)^m}{(m+1)!} (ad\ X)^m(Y) = g(ad\ X)(Y)
$$

as we wanted to prove.

Corollary 2. If F is a matrix function of t, then

$$
\frac{d}{dt} e^{F(t)} = e^{F(t)} g(ad\ F)(F'(t)).
$$

Proof. The map $t \longmapsto e^{F(t)}$ is the composition of $t \longmapsto F(t)$ and $X \longmapsto e^X$. The formula is then immediate from the chain rule for mappings.

Problems

1. Let $r > 0$ and let A_r be the algebra of power series converging on the disk $|z| < r$. Let V be a Banach space. For each $f = \sum f_n z^n \in A_r$ and any bounded operator $L(V \longrightarrow V)$ show that $f(L) = \sum_n f_n L^n$ is a well defined bounded operator provided $||L|| < r$ and that the map $f \longmapsto f(L)$ is an algebra homomorphism.
2. Let S be the space of skew hermitian matrices of order n. Prove that the polar decomposition map

$$
u, X \longmapsto u \exp X
$$

from $\mathrm{U}(n) \times S$ into $GL(n, \mathbf{C})$ is an analytic diffeomorphism with $\mathrm{GL}(n, \mathbf{C})$. Prove also the corresponding results for $\mathrm{SL}(n)$ and for $\mathrm{GL}(n, \mathbf{R})$, either directly or as a consequence of the result over $\mathbf{C}$.

7.3. Analytic derivation of BCH formula. We write

$$e^{uX}e^{vY} = e^{Z(u,v,X,Y)}$$

where Z is analytic in u, v near $(0,0)$ for each pair (X,Y). Moreover Z vanishes at $u = v = 0$. Differentiating with respect to v we get

$$e^{uX}e^{vY}Y = e^{Z}g(ad\ Z)\left(\frac{\partial Z}{\partial v}\right).$$

To calculate $\frac{\partial Z}{\partial v}$ we must invert $g(ad\ Z)$. Let $h = 1/g$. Then h is analytic for $|z| < 2\pi$ and $h(-z) = h(z) - \frac{1}{2}z$. Let us therefore set

$$f(z) = h(z) - \frac{1}{2}z = f(-z).$$

Then

$$f(z) = 1 + \sum_{p\geq 1} k_{2p}z^{2p}$$

where the k_r are rational numbers. Thus

$$\frac{\partial Z}{\partial v} = h(ad\ Z)(Y) = f(ad\ Z)(Y) + \frac{1}{2}[Z, Y].$$

To get $\frac{\partial Z}{\partial u}$ we invert the basic equation and use the evenness of f to derive from

$$e^{-vY}e^{-uX} = e^{-Z}$$

the equation

$$\frac{\partial Z}{\partial u} = h(ad\ Z)(Y) = f(ad\ Z)(X) - \frac{1}{2}[Z, X].$$

If we now set $u = v = t$ and $F(t, X, Y) = Z(t, t, X, Y)$, we get, from

$$\frac{dF}{dt} = \left(\frac{\partial Z}{\partial u} + \frac{\partial Z}{\partial v}\right)_{u=v=t}$$

the differential equation

$$\frac{dF}{dt} = f(ad\ F)(X + Y) + \frac{1}{2}[X - Y, F].$$

We write

$$F = z_1 t + z_2 t^2 + \dots \qquad ad\ F = tad\ z_1 + t^2 ad\ z_2 + \dots$$

so that

$$(ad\ F)^m = \sum_{n \geq 1} t^n \sum_{r_1 + \ldots + r_m = n} ad\ z_{r_1} \ldots ad\ z_{r_m}.$$

The differential equation is

$$\frac{dF}{dt} = X + Y + \sum k_{2p} (ad\ F)^{2p} (X + Y) + \frac{1}{2}[X - Y, F]$$

leading to the recursion formulae:

$$\begin{aligned} z_1 &= X + Y \\ (n+1) z_{n+1} &= \frac{1}{2}[X - Y, z_n] \\ &\quad + \sum_{p \geq 1, 2p \leq n} k_{2p} \sum_{r_i \geq 1, r_1 + \ldots + r_{2p} = n} [z_{r_1}, [z_{r_2}, \ldots [z_{r_{2p}}, X + Y]] \ldots]. \end{aligned}$$

Let

$$\mathcal{L}_n := \{ \text{ linear span of } n\text{-fold commutators involving } X, Y \}.$$

Then it is immediate by induction on n that

$$z_n \in \mathcal{L}_n \text{ for all } n \geq 2.$$

We have thus proved the following theorem.

Theorem 2. *There are unique homogeneous polynomial maps*

$$z_n : \mathfrak{g} \times \mathfrak{g} \longrightarrow \mathfrak{g}$$

with $z_1 = X + Y, z_n(X, Y) \in \mathcal{L}_n$ *for all* $n \geq 2$ *such that*

$$\exp X \exp Y = \exp Z(X, Y) \qquad Z(X, Y) = X + Y + \sum_{n \geq 2} z_n(X, Y),$$

the series converging for $|X|, |Y|$ *small.*

The convergence of the series for Z can be established independently by the so-called *method of majorants.* The idea is the following: the coefficients z_n are determined by a recursion scheme involving the coefficients k_{2p}. We can set up a parallel recursion scheme where the k_{2p} are replaced by $|k_{2p}|$. The solution to this parallel problem is always positive and defines a convergent solution by inspection, while the original solution is majorized by the solution to the parallel problem. This will prove the convergence.

Let

$$H(z) = 1 + \sum_{p \geq 1} |k_{2p}| z^{2p}$$

and let us consider the scalar initial value problem

$$\frac{dy}{dz} = \frac{1}{2}y + H(y), \qquad y(0) = 0.$$

Then there is a solution analytic in $|z| < \delta$; if we write

$$y = c_1 z + c_2 z^2 + \dots$$

then $c_1 = 1$ and

$$(n+1)c_{n+1} = \frac{1}{2}c_n + \sum_{p \geq 1, 2p \leq n} |k_{2p}| \sum_{r_i \geq 1, r_1 + \dots + r_{2p} = n} c_{r_1} \dots c_{r_{2p}}.$$

It is clear that all the c_n are ≥ 0. It is possible to show that the z_n are majorized by the c_n up to an exponential factor.

Problems

1. Let $M > 0$ be such that $|[A, B]| \leq M|A||B|$ for all matrices A, B. Let $\alpha = \max(|X|, |Y|)$. Prove the estimate

$$|z_n| \leq M^{n-1}(2\alpha)^n c_n$$

where c_n are as above. (*Hint*: Use induction on n.)

2. Let

$$\mathfrak{n} = \left\{ X \in \mathfrak{g} \;\middle|\; |X| < \frac{\delta}{2M} \right\}.$$

Prove that the BCH series converges absolutely for $X, Y \in \mathfrak{n}$.

7.4. Formal aspects of the BCH series and Dynkin's formula. At the formal level the BCH formula works with two non-commuting indeterminates x, y and the algebra $\mathcal{P}$ of formal power series in x and y with coefficients in a field k of characteristic 0. For any $f \in \mathcal{P}$ with zero constant term, the exponential e^f is defined by

$$e^f = 1 + f + \frac{f^2}{2!} + \dots.$$

Since f^n begins only with terms of order n it is immediate that this is an element of $\mathcal{P}$. Similarly

$$\log(1+f) = f - \frac{f^2}{2} + \frac{f^3}{3} - \dots.$$

Thus, taking $f = e^x e^y - 1$ we see that for some $z \in \mathcal{P}$ with zero constant term we have

$$e^x e^y = e^z.$$

If $\mathcal{P}_n$ is the subspace of $\mathcal{P}$ of polynomials in x and y which are homogeneous of degree n we can write

$$z = \sum_{n \geq 1} z_n, \qquad z_n \in \mathcal{P}_n.$$

There is a more refined grading of $\mathcal{P}_{rs}$; let $\mathcal{P}_{rs}$ be the subspace of $\mathcal{P}$ of elements which are separately homogeneous of degrees r in x and s in y. Then

$$\mathcal{P}_n = \oplus_{r+s=n} \mathcal{P}_{rs}.$$

In particular

$$z_n = \sum_{r+s=n} z_{rs}, \qquad z_{rs} \in \mathcal{P}_{rs}.$$

One can obtain an explicit formula for the z_{rs}. Notice that

$$e^x e^y = 1 + u, \qquad u = \sum_{p+q \geq 1} \frac{x^p y^q}{p!q!}$$

and so

$$z = \log(1+u) = \sum_{m\geq 1} \frac{(-1)^{m-1}}{m} \left(\sum_{p+q\geq 1} \frac{x^p y^q}{p!q!} \right)^m .$$

It follows from this that

$$z_{rs} = \sum_{m\geq 1} \frac{(-1)^{m-1}}{m} \sum_{p_i+q_i\geq 1, p_1+\ldots+p_m=r, q_1+\ldots+q_m=s} \frac{x^{p_1}y^{q_1}\ldots x^{p_m}y^{q_m}}{p_1!q_1!\ldots p_m!q_m!}.$$

Unfortunately this expression hides the fact that the z_{rs} are Lie elements. Let $\mathcal{F}_n$ be the linear span of commutators $[a_1,[a_2,\ldots,[a_{n-1},a_n]]]\ldots]$. Then one can show that $z_{rs}\in\mathcal{F}_{r+s}$. Moreover there is a remarkable explicit formula for the z_{rs} due to Dynkin that makes the fact that z_{rs} belongs to $\mathcal{F}_{rs}$ manifest. I shall sketch a proof of these results. However we need to use some sophisticated results from Lie algebra theory.

There are two issues: the first is to show that z_n is a Lie element, and the second is to find an explicit expression. Both of these are done in Serre's Notes. I shall describe in the exercises a different method for the first; in fact, the analytic method given earlier generalizes very nicely in the formal situation and gives a proof that the z_n are Lie elements. In the formal development we must replace the derivatives $\partial/\partial u, \partial/\partial v$ by suitable *continuous* derivations of $\mathcal{P}$; here continuity is with respect to the usual *adic* topology in which the ideals $\mathcal{M}_n$ of elements beginning with terms of degree n form a basis at 0.

Dynkin's formula. The question is to calculate z_{rs} explicitly given that they are Lie elements. This depends on the theory of *free Lie algebras* and their *universal enveloping algebras.* Let $\mathcal{A}$ be the free associative algebra generated by n elements $X_1, X_2, \ldots X_n$. We define $[u,v]=uv-vu$ for $u,v\in\mathcal{A}$ so that $\mathcal{A}$ becomes a Lie algebra. Let $\mathcal{L}$ the Lie subalgebra generated by (X_i). $\mathcal{A}$ is graded in the obvious manner. We write $\mathcal{A}_m$ for the graded components of $\mathcal{A}$. We have an endomorphism ψ of $\mathcal{A}$ into itself defined by

$$\psi(X_{i_1}\ldots X_{i_m}) = [X_{i_1},[X_{i_2},[\ldots[X_{i_{m-1}},X_{i_m}]]\ldots].$$

These considerations apply when $n=2$ and $X_1=x, X_2=y$. The key is the following result.

Lemma. *Let $\mathcal{L}_m=\mathcal{A}_m\cap\mathcal{L}$. Then $\mathcal{L}=\oplus\mathcal{L}_m$ and $\psi(u)=mu$ for $u\in\mathcal{L}_m$.*

Dynkin's formula is immediate. Since we know that $z_{rs} \in \mathcal{L}_{r+s}$ we have

$$(r+s)z_{rs} = \psi(z_{rs}).$$

In computing $\psi(x^{p_1}y^{q_1}\dots x^{p_m}y^{q_m})$ we note that it must vanish if $q_m \geq 2$ or if $q_m = 0$ and $p_m \geq 2$. Hence

$$z_{rs} = z'_{rs} + z''_{rs}$$

where $(r+s)z'_{rs}$ is

$$\sum_{m\geq 1} \frac{(-1)^{m-1}}{m} \sum_{p_i+q_i\geq 1, \sum p_i = r, q_1+\ldots+q_{m-1}=s-1} \frac{\mathrm{ad}x^{p_1}\mathrm{ad}y^{q_1}\dots\mathrm{ad}x^{p_m}(y)}{p_1!q_1!\dots p_m!}$$

and $(r+s)z''_{rs}$ is

$$\sum_{m\geq 1} \frac{(-1)^{m-1}}{m} \sum_{p_i+q_i\geq 1, p_1+\ldots+p_{m-1}=r-1, \sum q_i = s} \frac{\mathrm{ad}x^{p_1}\mathrm{ad}y^{q_1}\dots\mathrm{ad}y^{q_{m-1}}(y)}{p_1!q_1!\dots q_{m-1}!}.$$

Problems

1. (a) Prove that there are unique continuous endomorphisms D_1, D_2 of $\mathcal{P}$ such that $D_1 u = ru$ for $u \in \mathcal{A}_{rs}$ and $D_2 v = sv$ for $v \in \mathcal{A}_{rs}$. Prove that they are derivations.

   (b) Prove that

$$D_1 z = -\frac{1}{2}[z,x] + x + \sum_{p\geq 1} k_{2p}\mathrm{ad}\ z^{2p}(x)$$
$$D_2 z = \frac{1}{2}[z,x] + y + \sum_{p\geq 1} k_{2p}\mathrm{ad}\ z^{2p}(y).$$

   (c) Deduce that the z_{rs} are Lie elements, $z_1 = x+y$, and that $z_n := \sum_{r+s=n} z_{rs}$ satisfy the recursion formulae obtained in the analytic treatment.

2. Prove the formulae for z'_{rs}, z''_{rs}.

7.5. Remarks on free Lie algebras and their universal enveloping algebras. We shall complete this discussion with some elucidating remarks on free Lie algebras and enveloping algebras.

Universal enveloping algebra of a Lie algebra. Let $\mathfrak{g}$ be a Lie algebra. By *an enveloping algebra of* $\mathfrak{g}$ we mean a pair $(\mathcal{A}, f)$ where $\mathcal{A}$ is an associative algebra (with unit) and f is a linear map $\mathfrak{g} \longrightarrow \mathcal{A}$ such that $f(\mathfrak{g})$ generates $\mathcal{A}$ as an associative algebra and

$$f([X,Y]) = f(X)f(Y) - f(Y)f(X).$$

In other words f converts the unfamiliar bracket in $\mathfrak{g}$ into the familiar commutator bracket of an associative algebra. There are of course many enveloping algebras and so we are interested in the biggest, which is the universal enveloping algebra. An enveloping algebra $(\mathcal{U}, g)$ is said to be *universal* if for any enveloping algebra $(\mathcal{A}, f)$ there is a homomorphism $h(\mathcal{U} \longrightarrow \mathcal{A})$ such that

$$h(g(X)) = f(X) \qquad (\text{ for all } X \in \mathfrak{g}).$$

Since the image of $\mathfrak{g}$ in any enveloping algebra generates it, it is clear that h is unique. For the usual reasons the universal enveloping algebra is unique up to a unique isomorphism. Its existence is easy to see also. et T be the tensor algebra over $\mathfrak{g}$ and let J be the two-sided ideal in T generated by all elements of the form

$$u_{X,Y} = X \otimes Y - Y \otimes X - [X,Y].$$

The $(T/J, \pi)$ is a universal enveloping algebra of $\mathfrak{g}$ where π is the natural map $T \longrightarrow T/J$.

It is not obvious that π is injective on $\mathfrak{g}$. Actually much more is true. For simplicity let $\mathfrak{g}$ be finite dimensional and let $(X_i)_{1\le i\le n}$ be a basis of it. Let $\bar{X}_i = \pi(X_i)$. We now observe that

$$\bar{X}_i\bar{X}_j = \bar{X}_j\bar{X}_i + [\bar{X}_i, \bar{X}_j]$$

from which it follows that any *monomial* in the $\bar{X}_i$ can be written as a linear combination of *standard monomials*

$$\bar{X}_1^{r_1}\bar{X}_2^{r_2}\dots\bar{X}_n^{r_n}.$$

The fundamental result in this circle of ideas is

Theorem (Poincaré-Birkhoff-Witt). *The standard monomials form a basis for the universal enveloping algebra. In particular* $\mathfrak{g}$ *imbeds into the universal enveloping algebra.*

In a similar spirit we shall define what we mean by the *free Lie algebra generated by the* X_i. It is by definition a Lie algebra $\mathcal{L}$ containing the X_i and generated by it, with the following *universal property*: if $\mathfrak{m}$ is any Lie algebra and $X_i' \in \mathfrak{m}$ are arbitrary elements, there is a (unique) Lie morphism $f(\mathcal{L} \longrightarrow \mathfrak{m})$ such that $f(X_i) = X_i'$. The free Lie algebra, if it exists, is unique. Its existence is given by

Theorem. *Let* $\mathcal{A}$ *be the free associative algebra generated by the* X_i *and let* $\mathcal{L}$ *the Lie subalgebra generated by the* X_i. *Then* $\mathcal{L}$ *is the free Lie algebra generated by the* X_i.

Proof. This is an immediate consequence of the existence of the universal enveloping algebra and the PBW theorem. Let $\mathfrak{m}$ be any Lie algebra and $X_i' \in \mathfrak{m}$ be arbitrary elements. Let $\mathcal{M}$ be the universal enveloping algebra of $\mathfrak{m}$ with $\mathfrak{m} \subset \mathcal{M}$. There is a homomorphism $h(\mathcal{A} \longrightarrow \mathcal{M})$ such that $h(X_i) = X_i'$. The restriction of h to $\mathcal{L}$ is a Lie algebra map. But we must show that this restriction maps $\mathcal{L}$ into $\mathfrak{m}$. Let $\mathcal{L}'$ be the preimage of $\mathfrak{m}$ by h. Then $\mathcal{L}'$is a Lie algebra and contains all the X_i, hence contains $\mathcal{L}$, proving that $h(\mathcal{L}) \subset \mathfrak{m}$.

We are now in a position to address the issues concerning the map ψ occurring in the proof of Dynkin's formula. We consider first the *adjoint representation* of $\mathcal{L}$ acting on $\mathcal{A}$ by

$$\text{ad } x : u \longmapsto xu - ux \qquad (u \in \mathcal{A}).$$

We may view ad as a map of $\mathcal{L}$ into $\text{End}(\mathcal{A})$ such that

$$\text{ad } ([x, y]) = \text{ad } x\text{ad } y - \text{ad } y\text{ad } x$$

and so it extends to a homomorphisn θ of $\mathcal{A}$ into $\text{End}(\mathcal{A})$:

$$\theta(uv) = \theta(u)\theta(v) \quad (u, v \in \mathcal{A}), \qquad \theta(x) = \text{ad } x \quad (x \in \mathcal{L}).$$

Recall now that

$$\psi(X_{i_1} \dots X_{i_m}) = [X_{i_1}, [X_{i_2}, [\dots [X_{i_{m-1}}, X_{i_m}]] \dots].$$

It follows from this that

$$\psi(uv) = \theta(u)(\psi(v)).$$

First we shall show that the $\psi(X_{i_1} \dots X_{i_m})$ span $\mathcal{L}$. Indeed, if the span is $\mathcal{L}' \subset \mathcal{L}$, then $\mathcal{L}'$ is stable under ad X_i for all i, and hence under $\theta(u)$ for all $u \in \mathcal{A}$. In particular $[\mathcal{L}, \mathcal{L}'] \subset \mathcal{L}'$, hence $[\mathcal{L}', \mathcal{L}'] \subset \mathcal{L}'$, or $\mathcal{L}'$ is a Lie algebra. Since it contains all the X_i it contains $\mathcal{L}$, hence must be equal to $\mathcal{L}$. Now $\mathcal{A}$ is naturally graded and the above result shows that $\mathcal{L}$ is spanned by homogeneous elements. Let

$$\mathcal{L}_m = \mathcal{L} \cap \mathcal{A}_m.$$

Then

$$\mathcal{L} = \oplus_m \mathcal{L}_m, \quad \mathcal{L}_m = \text{span of } \psi(X_{i_1} \dots X_{i_m}).$$

We shall now prove by induction on m that

$$\psi(u) = mu \quad (u \in \mathcal{L}_m).$$

The result is trivial for $m = 1$. let $m > 1$ and let us assume the result for lower values of m. Let $u = [X_i, v]$ where $v = \psi(w)$ for some $w \in \mathcal{A}_{m-1}$. Then

$$\begin{aligned} \psi(u) = \psi(X_i v) - \psi(v X_i) &= [X_i, \psi(v)] - \theta(v)(X_i) \\ &= (m-1)[X_i, v] - [v, X_i] \end{aligned}$$

because $\theta(v) = \text{ad } v$ (since $v \in \mathcal{L}$). hence

$$\psi(u) = (m-1)[X_i, v] + [X_i, v] = m[X_i, v] = mu.$$