Proof of the PBW theorem

Theorem. - [Poincaré - Birkhoff - Witt] - Let $\mathfrak{g}$ be a Lie algebra over a field k. Let $\{x_i\}_{i\in I}$ be a basis of $\mathfrak{g}$ and let $(I, <)$ be a total ordering on the indexing set I. Then, the set of ordered monomials

$$\mathfrak{X} = \left\{ x_{i_1} \cdots x_{i_r} \ : \ r \in \mathbb{Z}_{\geq 0},\ i_1 \leq i_2 \leq \ldots \leq i_r \text{ in } I \right\}$$

(for $r=0$, the empty word is, by convention, 1)

forms a basis of $\mathcal{U}(\mathfrak{g})$.

Proof. - We begin by showing that the set $\mathfrak{X}$ spans $\mathcal{U}(\mathfrak{g})$. Let us argue by induction on the filter-degree. Recall that the canonical filtration on $\mathcal{U}(\mathfrak{g})$ is given as follows: $F_0 = k$ and

$$F_n = \text{Image of } \left(\bigoplus_{i=0}^{n} T^i(\mathfrak{g}) \hookrightarrow T^{\cdot}(\mathfrak{g}) \twoheadrightarrow \mathcal{U}(\mathfrak{g}) \right)$$

In other words, F_n = linear (over k) span of monomials

$$\left\{ \begin{array}{l} x_{j_1} \cdots x_{j_r} \ ; \ j_1, \ldots, j_r \in I \\ 0 \leq r \leq n \end{array} \right\}.$$

Claim. - F_n is spanned by ordered monomials

$$\mathfrak{X}_n = \left\{ x_{i_1} \cdots x_{i_r} \ : \ \begin{array}{l} 0 \leq r \leq n, \\ i_1 \leq i_2 \leq \ldots \leq i_r \text{ in } I \end{array} \right\}$$

Proof. - Clear for $n=0$, $\mathfrak{X}_0 = \{1\}$ basis of $F_0 = k$.

Assume $\mathfrak{X}_{n-1}$ forms a spanning set of F_{n-1}.

For an arbitrary monomial $x_{j_1} \cdots x_{j_n}$ - let $\sigma \in S_n$ be such that $i_1 = j_{\sigma(1)} \leq i_2 = j_{\sigma(2)} \leq \ldots \leq i_n = j_{\sigma(n)}$

Recall that we proved in Lecture 5 §3 that

$$x_{j_1} \cdots x_{j_n} - x_{j_{\sigma(1)}} \cdots x_{j_{\sigma(n)}} \in F_{n-1}$$

Thus, by the induction hypothesis, $x_{j_1} \cdots x_{j_n} \in$ Span of $\mathcal{X}_n$, and therefore $\mathcal{X}_n$ spans F_n. The claim is proved.

The proof of the linear independence of $\mathcal{X}$ relies on the following

Lemma - Let $P = k[z_i : i \in I]$ be the polynomial ring in variables $\{z_i\}_{i \in I}$. Then, $\exists$ a unique representation of $\mathcal{X}$ on P satisfying the following conditions:

(*) -
I. For every $i \in I$ and a monomial $z_{i_1} \cdots z_{i_\ell}$ in P
$$x_i \cdot (z_{i_1} \cdots z_{i_\ell}) = z_i z_{i_1} \cdots z_{i_\ell} \text{ modulo terms of degree} \leq \ell$$
If $i \leq i_1 \leq \cdots \leq i_\ell$, then $x_i \cdot (z_{i_1} \cdots z_{i_\ell}) = z_i z_{i_1} \cdots z_{i_\ell}$

Given the lemma, the proof of linear independence of $\mathcal{X}$ proceeds as follows. Let $\sum_{\underline{i} \in \mathcal{X}} c_{\underline{i}} x(\underline{i}) = 0$ be a finite linear dependence relation.

Here, $x(\underline{i}) = x_{i_1} \cdots x_{i_\ell}$ if $\underline{i} = (i_1 \leq \cdots \leq i_\ell) \in \mathcal{X}$. Note that under the alg. hom. $\rho : \mathcal{U}(\mathfrak{g}) \to \mathrm{End}(P)$, $\rho(x(\underline{i})) \cdot 1 = z_{i_1} \cdots z_{i_\ell}$

Thus $\rho\left(\sum_{\underline{i} \in \mathcal{X}} c_{\underline{i}} x(\underline{i})\right) \cdot 1 = \sum_{\underline{i} \in \mathcal{X}} c_{\underline{i}} z_{i_1} \cdots z_{i_\ell} = 0$ implies

all $c_{\underline{i}}$'s are zero - since $\{z_{i_1} \cdots z_{i_\ell}\}_{i_1 \le \ldots \le i_\ell \text{ in } I,\ \ell \ge 0}$ is a basis of $P = k[z_i : i \in I]$.

□

Proof of the lemma - Consider the filtration on P: $P_0 = k$

P_n = poly's of degree $\le n$. Note that P_n is a vector subspace of P with basis $\left\{ z_{i_1} \cdots z_{i_\ell} : \begin{matrix} 0 \le \ell \le n \\ i_1 \le \ldots \le i_\ell \text{ in } I \end{matrix} \right\}$

Our aim is to define $\mathfrak{g}$-action on P satisfying the condition (*) on the previous page. We will do it inductively on n.

i.e., construct $\rho_n : \mathfrak{g} \to \underset{\text{v.s.}}{\mathrm{Hom}}(P_n, P_{n+1})$ so that

$(*)_n$
$$\rho_n(x_i) \cdot z_{i_1} \cdots z_{i_\ell} = z_i z_{i_1} \cdots z_{i_\ell} \mod P_\ell$$
$$= z_i z_{i_1} \cdots z_{i_\ell} \text{ if } i \le i_1 \le \ldots \le i_\ell.$$

for all $0 \le \ell \le n$, $i_1 \le \ldots \le i_\ell$ in I.

and $\rho_n \big|_{P_{n-1}} = \rho_{n-1}$ (this will follow from the uniqueness)

We will set (at the end) $\rho(x) : P \to P$ $(\forall\, x \in \mathfrak{g})$ the unique linear map whose restriction to P_n is $\rho_n(x)$.

For this to be a Lie alg. hom, we will have to check

$(**)_n$
$$\rho_n(x_i)\rho_n(x_j) - \rho_n(x_j)\rho_n(x_i) - \rho_n([x_i, x_j])$$
$$= 0 \quad \text{on } P_{n-1}.$$

$n=0$. By $(*)_0$, $\rho_0(x_i)\cdot 1 = z_i \quad \forall\, i\in I$.

$(**)_0$ is vacuous. $\rho_0(x): P_0 = k \longrightarrow P_1$, $x \in g$

is thus uniquely defined.

Assume $\rho_0, \rho_1, \dots, \rho_n$ have been defined in a consistent manner - i.e. $\rho_n\big|_{P_\ell} = \rho_\ell \quad \forall\, 0\le \ell \le n$ - and $(**)_n$ holds.

$$\rho_{n+1}(x_i)\cdot z_{i_1}\cdots z_{i_\ell} \;:=\; \rho_n(x_i)\cdot z_{i_1}\cdots z_{i_\ell} \quad \text{if } \ell < n+1 \text{ (forced by consistency)}$$
$$i_1\le \dots \le i_\ell$$

If $\ell = n+1$, we have two cases to consider

$i \le i_1$: $\rho_{n+1}(x_i)\cdot z_{i_1}\cdots z_{i_{n+1}} := z_i\, z_{i_1}\cdots z_{i_{n+1}}$ forced by $(*)_{n+1}$

$i > i_1$ $\quad \rho_{n+1}(x_i)\cdot z_{i_1}\cdots z_{i_{n+1}} = \underbrace{\rho_{n+1}(x_i)\,\rho_{n+1}(x_{i_1})}\cdot z_{i_2}\cdots z_{i_{n+1}}$

Since $\rho_{n+1}(x_i)\cdot z_{i_2}\cdots z_{i_{n+1}}$ has been defined, and

by $(*)_n$ $\quad = z_i\, z_{i_2}\cdots z_{i_{n+1}} + v$, where $v\in P_n$

we are forced to set (by $(**)$ condition)

$$\rho_{n+1}(x_i)\cdot z_{i_1}\cdots z_{i_{n+1}} = \rho_{n+1}(x_{i_1})\cdot\rho_{n+1}(x_i)\cdot z_{i_2}\cdots z_{i_{n+1}}$$
$$+\ \rho_{n+1}\big([x_i, x_{i_1}]\big)\cdot z_{i_2}\cdots z_{i_{n+1}}$$
$$\rho_{n+1}(x_{i_1})\cdot(z_i z_{i_2}\cdots z_{i_{n+1}})$$
$$= \boxed{z_i\, z_{i_1}\cdots z_{i_{n+1}}} + \rho_{n+1}(x_{i_1})\cdot v + \rho_{n+1}\big([x_i, x_{i_1}]\big)\cdot z_{i_2}\cdots z_{i_{n+1}}$$

Thus there is a unique way to extend $\rho_n(x) : P_n \to P_{n+1}$

to $\rho_{n+1}(x) : P_{n+1} \to P_{n+2}$ so that $(*)_{n+1}$ and $(**)_{n+1}$ hold,

namely:

$$\rho_{n+1}(x_i)\cdot z_{i_1}\cdots z_{i_\ell} = \begin{cases} \rho_n(x_i)\cdot z_{i_1}\cdots z_{i_\ell}; & \text{if } 0\le \ell \le n \\ z_i z_{i_1}\cdots z_{i_{n+1}}; & \text{if } \ell = n+1 \text{ and } i\le i_1\le i_2\le\ldots\le i_{n+1} \\ z_i z_{i_1}\cdots z_{i_{n+1}} + \rho_{n+1}(x_{i_1})\cdot v; & \text{if } \ell = n+1,\ i > i_1, \text{ and } \rho_{n+1}(x_i)\cdot z_{i_2}\cdots z_{i_{n+1}} = z_i z_{i_2}\cdots z_{i_{n+1}} + v,\ v\in P_n \\ + \rho_{n+1}\big([x_i, x_{i_1}]\big)\cdot z_{i_2}\cdots z_{i_{n+1}} & \end{cases}$$

Note that $(*)_{n+1}$ is true by the formulae above. Moreover,

$(**)_{n+1}$:

$$\Big(\rho_{n+1}(x_i)\circ\rho_{n+1}(x_j) - \rho_{n+1}(x_j)\circ\rho_{n+1}(x_i) - \rho_{n+1}\big([x_i,x_j]\big)\Big)\cdot z_{K_1}\cdots z_{K_\ell}$$

$= 0$ holds for all $0\le \ell \le n-1$

By construction, it also holds when $\ell = n$, $K_1 \le \ldots \le K_\ell$ and one (or both) of i, j is smaller than K_1 as we see below:

e.g. if $\begin{pmatrix} i \le K_1 \\ \& \\ i < j \end{pmatrix}$ [for notational ease, I will write $x\cdot\xi$ for $\rho_{n+1}(x)(\xi)$, $x\in g$, $\xi\in P_{n+1}$]

then $x_j\cdot(x_i\cdot z_{K_1}\cdots z_{K_n}) = x_j\cdot(z_i z_{K_1}\cdots z_{K_n})$

Since $j > i$, by defn. above this equals

$$x_i\cdot(x_j\cdot z_{K_1}\cdots z_{K_n}) + [x_i, x_j]\cdot z_{K_1}\cdots z_{K_n}$$

Thus, the only remaining case is when both $i, j > K_1$.

In this case, $x_i \cdot (x_j \cdot z_{K_1} \cdots z_{K_n}) = x_i \cdot (x_j \cdot (x_{K_1} \cdot z_{K_2} \cdots z_{K_n}))$

$= x_i \cdot \left(x_{K_1} \cdot (x_j \cdot z_{K_2} \cdots z_{K_n}) + [x_j, x_{K_1}] \cdot z_{K_2} \cdots z_{K_n} \right)$

Now $x_j \cdot z_{K_2} \cdots z_{K_n} = z_j z_{K_2} \cdots z_{K_n} + u, \quad u \in P_{n-1}$

and $(**)$ holds for x_i, x_{K_1} acting on $z_j z_{K_2} \cdots z_{K_n}$
since $K_1 \leq j, K_2, \ldots, K_n$

and on u, since $u \in P_{n-1}$.

Thus, $x_i \cdot (x_{K_1} \cdot (x_j \cdot z_{K_2} \cdots z_{K_n}))$

$= x_{K_1} \cdot (x_i \cdot (x_j \cdot z_{K_2} \cdots z_{K_n})) + [x_i, x_{K_1}] \cdot (x_j \cdot z_{K_2} \cdots z_{K_n})$

$\Rightarrow \quad x_i \cdot (x_j \cdot z_{K_1} \cdots z_{K_n}) = x_{K_1} \cdot (x_i \cdot (x_j \cdot z_{K_2} \cdots z_{K_n}))$

$+ [x_i, x_{K_1}] \cdot (x_j \cdot z_{K_2} \cdots z_{K_n})$

$+ [x_j, x_{K_1}] \cdot (x_i \cdot z_{K_2} \cdots z_{K_n})$

$+ [x_i, [x_j, x_{K_1}]] \cdot z_{K_2} \cdots z_{K_n}$

Switching the role of i, j and subtracting, we get

$(x_i x_j - x_j x_i) \cdot z_{K_1} \cdots z_{K_n} = x_{K_1} [x_i, x_j] \cdot z_{K_2} \cdots z_{K_n}$

$+ \left([x_i, [x_j, x_{K_1}]] - [x_j [x_i, x_{K_1}]] \right) \cdot z_{K_2} \cdots z_{K_n}$

$= \left(x_{K_1} [x_i, x_j] + [[x_i, x_j], x_{K_1}] \right) \cdot z_{K_2} \cdots z_{K_n}$

by Jacobi id.

$= [x_i, x_j] \cdot x_{K_1} \cdot z_{K_2} \cdots z_{K_n} = [x_i, x_j] \cdot z_{K_1} \cdots z_{K_n}$

as desired. □