Proof of the Poincaré-Birkhoff-Witt Theorem

Notation so far

L a finite dimensional Lie algebra over k.

$T(L) = \bigoplus_{m\geq 0} \bigotimes^m L$ – the tensor algebra.

$S(L) = T(L)/I$, I the ideal generated by $\{x\otimes y - y\otimes x \mid x,y \in L\}$ – the symmetric algebra.

$U(L) = T(L)/J$, J the ideal generated by $\{x\otimes y - y\otimes x - [xy] \mid x,y\in L\}$ – the universal enveloping algebra.

$\pi : T(L) \to U(L)$ – the quotient map.

Let $T_n = \bigoplus_{m=0}^n \bigotimes^m L$. Let S_n, U_n denote the image of T_n in $S(L), U(L)$ respectively. Then, $U_0 < U_1 < \dots$ is a filtration of $U(L)$.

Let $U^m = U_m/U_{m-1}$; $grU = \bigoplus_{m\geq 0} U^m$ is the associated graded algebra.

$\phi : T(L) \to grU$ is induced from the maps $\bigotimes^m L \to U_m \to U^m = U_m/U_{m-1}$ for each m. Since $\phi(I) = 0$, ϕ induces a surjective algebra homomorphism $\theta : S(L) \to grU$.

The PBW Theorem *$\theta : S(L) \to grU$ is an isomorphism.*

To prove it, we need a technical lemma. Let $x_1, \dots, x_n$ be a fixed basis for L. Let $I(m)$ denote all m-tuples $\mathbf{i} = (i_1, \dots, i_m)$ with $1 \leq i_j \leq n$. Let $x_{\mathbf{i}}$ be the element $x_{i_1} \otimes \cdots \otimes x_{i_m} \in T(L)$, and $\bar{x}_{\mathbf{i}}$ be its image in $S(L)$. Say $\mathbf{i}$ is increasing if $i_1 \leq \cdots \leq i_m$. Then, $\{x_{\mathbf{i}} \mid m \in \mathbb{Z}_{\geq 0}, \mathbf{i} \in I(m)\}$ is a basis for $T(L)$ and $\{\bar{x}_{\mathbf{i}} \mid m \in \mathbb{Z}_{\geq 0}, \mathbf{i} \in I(m), \mathbf{i} \text{ increasing}\}$ is a basis for $S(L)$. Write $i \leq \mathbf{j}$ if $i \leq j_s$ for all $1 \leq s \leq m$.

Technical Lemma *For each $m \in \mathbb{Z}_{\geq 0}$ there exists a unique linear map $f_m : L\otimes S_m \to S(L)$ satisfying*

(i) $f_m(x_i \otimes \bar{x}_{\mathbf{j}}) = \bar{x}_i\bar{x}_{\mathbf{j}}$ for $i \leq \mathbf{j}, \mathbf{j} \in I(m)$;

(ii) $f_m(x_i \otimes \bar{x}_{\mathbf{j}}) - \bar{x}_i\bar{x}_{\mathbf{j}} \in S_k$ for $k \leq m, \mathbf{j} \in I(k)$.

(iii) $f_m(x_i \otimes f_m(x_j \otimes \bar{x}_{\mathbf{k}})) = f_m(x_j \otimes f_m(x_i \otimes \bar{x}_{\mathbf{k}})) + f_m([x_ix_j] \otimes \bar{x}_{\mathbf{k}})$ for all $\mathbf{k} \in I(m-1)$.

Moreover, the restriction of f_m to $L \otimes S_{m-1}$ equals f_{m-1}.

(The expression in (iii) makes sense given (ii).)

PROOF. First, note that the restriction of f_m to $L \otimes S_{m-1}$ automatically satisfies (i)–(iii), so it must coincide with f_{m-1} by the asserted uniqueness. To verify existence and uniqueness, we proceed by induction on m. For $m = 0$, $f_0(x_i \otimes 1) = \bar{x}_i$ is the only possibility, and the induction starts. Now suppose by induction that a unique f_{m-1} satisfying (i)–(iii) has been constructed. We wish to extend f_{m-1} to a map f_m; for this, it will suffice to define $f_m(x_i \otimes \bar{x}_{\mathbf{j}})$ when $\mathbf{j} \in I(m)$ is increasing.

If $i \leq \mathbf{j}$, (i) cannot hold unless we define $f_m(x_i \otimes \bar{x}_{\mathbf{j}})$ to be $\bar{x}_i\bar{x}_{\mathbf{j}}$. If $i \leq \mathbf{j}$ fails, then $j_1 < i$. Let $j = j_1, \mathbf{k} = (j_2, \dots, j_m) \in I(m-1)$. Then, by (i) and the induction hypothesis, $\bar{x}_{\mathbf{j}} = \bar{x}_j\bar{x}_{\mathbf{k}} = f_{m-1}(x_j \otimes \bar{x}_{\mathbf{k}})$. Since $j \leq \mathbf{k}$, $f_m(x_j \otimes \bar{x}_{\mathbf{k}}) = \bar{x}_{\mathbf{j}}$ is already defined, so the left hand side of (iii) becomes $f_m(x_i \otimes \bar{x}_{\mathbf{j}})$, which is what we are trying to define. By induction, $f_m(x_i \otimes \bar{x}_{\mathbf{k}}) = f_{m-1}(x_t \otimes \bar{x}_{\mathbf{k}})$ is already known; it equals

$\bar{x}_i\bar{x}_{\mathbf{k}} + y$ for some $y \in S_{m-1}$. So, the right hand side of (iii) is already defined to be

$$\bar{x}_j\bar{x}_i\bar{x}_{\mathbf{k}} + f_{m-1}(x_j \otimes y) + f_{m-1}([x_ix_j] \otimes f_{\mathbf{k}}).$$

This shows that f_m can be defined, and in a unique way. Moreover, (i) and (ii) clearly hold. So, it suffices to show that (iii) holds. This is clear in case $j < i, j \leq \mathbf{k}$, hence (as $[x_jx_i] = -[x_ix_j]$) in case $i < j, i \leq \mathbf{k}$. If $i = j$, (iii) is obvious. So, it remains to consider the case where neither $i \leq \mathbf{k}, j \leq \mathbf{k}$ is true. Let $k = k_1, \mathbf{l} = (k_2, \ldots, k_{m-1}) \in I(m-2)$, where $k \leq \mathbf{l}, k < i, k < j$. Abbreviate $f_m(x \otimes \bar{x})$ by $x\bar{x}$ whenever $x \in L, \bar{x} \in S_m$.

By induction, $x_j\bar{x}_{\mathbf{k}} = x_j(x_k\bar{x}_{\mathbf{l}}) = x_k(x_j\bar{x}_{\mathbf{l}}) + [x_jx_k]\bar{x}_{\mathbf{l}}$ and $x_j\bar{x}_{\mathbf{l}} = \bar{x}_j\bar{x}_{\mathbf{l}} + w$ for some $w \in S_{m-2}$. Since $k \leq \mathbf{l}, k < j$, (iii) already holds for $x_i(x_k(\bar{x}_j\bar{x}_{\mathbf{l}}))$. By induction, (iii) applies to $x_i(x_kw)$ hence to $x_i(x_k(x_j\bar{x}_{\mathbf{l}}))$. Consequently

$$x_i(x_j\bar{x}_{\mathbf{k}}) = x_k(x_i(x_j\bar{x}_{\mathbf{l}})) + [x_ix_k](x_j\bar{x}_{\mathbf{l}}) + [x_jx_k](x_i\bar{x}_{\mathbf{l}}) + [x_i[x_jx_k]]\bar{x}_{\mathbf{l}}.$$

Now, interchange i, j in this expression and subtract the two resulting equations to obtain:

$$\begin{aligned} x_i(x_j\bar{x}_{\mathbf{k}}) - x_j(x_i\bar{x}_{\mathbf{k}}) &= x_k(x_i(x_j\bar{x}_{\mathbf{l}})) - x_k(x_j(x_i\bar{x}_{\mathbf{l}})) + [x_i[x_jx_k]]\bar{x}_{\mathbf{l}} - [x_j[x_ix_k]]\bar{x}_{\mathbf{l}} \\ &= x_k([x_ix_j]\bar{x}_{\mathbf{l}}) + [x_i[x_jx_k]]\bar{x}_{\mathbf{l}} + [x_j[x_kx_i]]\bar{x}_{\mathbf{l}} \\ &= [x_ix_j](x_k\bar{x}_{\mathbf{l}}) + ([x_k[x_ix_j]] + [x_i[x_jx_k]] + [x_j[x_kx_i]])\bar{x}_{\mathbf{l}} \\ &= [x_ix_j]\bar{x}_{\mathbf{k}}. \end{aligned}$$

This proves (iii) and the lemma. □

We can now prove the PBW Theorem. Let $t \in \bigotimes^m L$. We need to show that $\pi(t) \in U_{m-1}$ implies that $t \in I$, so that θ is indeed injective. So, suppose $\pi(t) \in U_{m-1}$. Then, there is some $t' \in T_{m-1}$ such that $\pi(t) = \pi(t')$, hence that $t - t' \in J$. Now, $t - t' \in T_m \cap J$ and the homogeneous component of degree m of $t - t'$ is t. So, the result follows if we can prove:

Claim *Let $t \in T_m \cap J$. The homogeneous component t_m of t of degree m lies in I.*

PROOF. By the technical lemma, we can define a linear map $f : L \otimes S(L) \to S(L)$ satisfying (i), (ii), (iii) for all m. Property (iii) then ensures that this makes $S(L)$ into an L-module, and property (ii) shows that

(†) $x_i.\bar{x}_{\mathbf{j}} \equiv \bar{x}_i\bar{x}_{\mathbf{j}}$ (modulo S_m) for $\mathbf{j} \in I(m)$.

So, $S(L)$ is a $T(L)$-module such that J acts as zero, by the universal property of $U(L)$. Consequently, t acts as zero on $S(L)$. So, $t.1 = 0$.

Now suppose $t_m = \sum_{\mathbf{i} \in I(m)} a_{\mathbf{i}}x_{\mathbf{i}}$ for $a_{\mathbf{i}} \in k$. Then, using (†), $t.1$ is a polynomial whose term of highest degree m is $\sum_{\mathbf{i} \in I(m)} a_{\mathbf{i}}\bar{x}_{\mathbf{i}}$. As $t.1 = 0$, this is zero in $S(L)$, which shows precisely that $t_m \in I$ as required. □

This completes the proof of the PBW theorem. It is also valid (with essentially the same proof) for L infinite.

Two results from linear algebra

Throughout, the base field is $\mathbb{C}$. A linear map $\theta : V \to V$ is *semisimple* if it is diagonalisable, or, equivalently, if all the roots of its minimal polynomial are distinct. It is *nilpotent* if $\theta^n = 0$ for some $n > 0$.

Theorem *Let V be a finite dimensional $\mathbb{C}$-space, $\theta \in \operatorname{End} V$. Then, there exist unique $\theta_s, \theta_n \in \operatorname{End} V$ such that*

(i) $\theta = \theta_s + \theta_n$;

(ii) θ_s is semisimple, θ_n is nilpotent;

(iii) θ_s and θ_n commute.

Moreover, θ_s and θ_n can each be written as polynomials in θ without constant term.

PROOF. Let $a_1, \ldots, a_k$ (with multiplicities $m_1, \ldots, m_k$) be the distinct eigenvalues of θ, so that the characteristic polynomial is $\prod (T - a_i)^{m_i}$. If $V_i = \ker(\theta - a_i.1)^{m_i}$, then V is the direct sum of the subspaces $V_1, \ldots, V_k$, each stable under θ. On V_i, θ has characteristic polynomial $(T - a_i)^{m_i}$. Now apply the Chinese remainder theorem for the ring $\mathbb{C}[T]$ to find a polynomial $p(T)$ such that

$$p(T) \equiv a_i \pmod{(T - a_i)^{m_i}}; \qquad p(T) \equiv 0 \pmod{T}.$$

Set $q(T) = T - p(T)$. Evidently, each of $p(T), q(T)$ have no constant term as $p(T) \equiv 0 \pmod{T}$.

Set $\theta_s = p(\theta), \theta_n = q(\theta)$. Since they are polynomials in θ, θ_s and θ_n commute with each other. They also stabilise all subspaces of V stabilised by θ, in particular, each V_i. The congruence $p(T) \equiv a_i \pmod{(T - a_i)^{m_i}}$ shows that the restriction of $\theta_s - a_i.1$ to V_i is zero for all i, hence that θ_s acts diagonally on V_i with single eigenvalue a_i. By definition, $\theta_n = \theta - \theta_s$ which makes it clear that θ_n is nilpotent.

It remains to prove the uniqueness assertion. Suppose $\theta = \theta'_s + \theta'_n$ is another such decomposition. So, $\theta_s - \theta'_s = \theta'_n - \theta_n$. All endomorphisms commute, and sums of semisimple (resp. nilpotent) endomorphisms are semisimple (resp. nilpotent). But, 0 is the only endomorphism that is both semisimple and nilpotent, forcing $0 = \theta_s - \theta'_s = \theta'_n - \theta_n$. □

The decomposition $\theta = \theta_s + \theta_n$ is known as the *Jordan decomposition* of θ, and θ_s, θ_n are known as the *semisimple* and *nilpotent* parts of θ respectively.

The second result from linear algebra is really a corollary of the Jordan decomposition. We have already used this – without giving a proof – when we proved Cartan's criterion.

Theorem *Let $A \subset B$ be two subspaces of $\mathfrak{gl}(V)$, V finite dimensional. Set $M = \{x \in \mathfrak{gl}(V) \mid [x, B] \subset A\}$. Suppose $x \in M$ satisfies $Trace(xy) = 0$ for all $y \in M$. Then, x is nilpotent.*

PROOF. Let $x = s + n$ ($s = x_s$, $n = x_n$) be the Jordan decomposition of x. Fix a basis $v_1, \ldots, v_m$ of v relative to which s has matrix $\operatorname{diag}(a_1, \ldots, a_m)$. Let E be the vector subspace of $\mathbb{C}$ over $\mathbb{Q}$ spanned by the eigenvalues $a_1, \ldots, a_m$. We have to show

that $s = 0$, or equivalently, that $E = 0$. Since E has finite dimension over $\mathbb{Q}$, it will suffice to show that the dual space E^* is zero, ie that any linear map $f : E \to \mathbb{Q}$ is zero.

Given f, let y be that element of $\mathfrak{gl}(V)$ whose matrix relative to our given basis is $\mathrm{diag}(f(a_1), \ldots, f(a_m))$. If $\{e_{ij}\}$ is the corresponding basis of $\mathfrak{gl}(V)$, check that $\mathrm{ad}\, s(e_{ij}) = (a_i - a_j)e_{ij}, \mathrm{ad}\, y(e_{ij}) = (f(a_i) - f(a_j))e_{ij}$. Now let $r(T) \in \mathbb{C}[T]$ be a polynomial without constant term satisfying $r(a_i - a_j) = f(a_i) - f(a_j)$ (which exists by Lagrange interpolation). So, $\mathrm{ad}\, y = r(\mathrm{ad}\, s)$.

Now, $\mathrm{ad}\, s$ is the semisimple part of $\mathrm{ad}\, x$, so by the Jordan decomposition, it can be written as a polynomial in $\mathrm{ad}\, x$ without constant term. By hypothesis, $\mathrm{ad}\, x$ maps B into A, so $\mathrm{ad}\, s$ does, and so $\mathrm{ad}\, y$ does too. So, $y \in M$. So by hypothesis, $Trace(xy) = 0$, so $\sum a_i f(a_i) = 0$. The left side is a $\mathbb{Q}$-linear combination of elements of E. Applying f, we obtain $\sum (f(a_i))^2 = 0$. Since the numbers $f(a_i)$ are rational, this forces them all to be zero. $\square$

Some corrections/omissions

1. Several people had trouble showing that if L is semisimple and V is an L-module, then $\mathrm{tr}_V\, x = 0$ for all $x \in L$. To prove this, note that $x \mapsto \mathrm{tr}_V\, x$ is a homomorphism of Lie algebras, so the kernel of trace is an ideal of L, with $L/\ker \mathrm{tr}_V$ an *abelian* Lie algebra. So, the kernel of trace contains the derived algebra L' of L. So, the result follows once we've checked that $L' = L$ if L is semisimple.

For this, note that L can be written as $L_1 \oplus \cdots \oplus L_t$ as a direct sum of simple ideals. Moreover, the ideal L' is a sum of some subset of these ideals (since any ideal of L is of this form). So, L/L' is isomorphic to the sum of the remaining simple ideals. But L/L' is abelian, whilst simple Lie algebras are not, so this implies that $L = L'$.

2. I've proved that if $x \in \mathfrak{gl}(V)$ is a nilpotent matrix, then $\mathrm{ad}\, x$ is nilpotent (using the binomial expansion). But in Theorem 4.5, I also assumed that if $x \in \mathfrak{gl}(V)$ is a diagonalisable matrix, then $\mathrm{ad}\, x$ is diagonalisable. I didn't prove this... So, pick a basis for V so that x is a diagonal matrix. Now check that $\mathrm{ad}\, x$ acts diagonally on the basis $e_{i,j}$, where $e_{i,j}$ is the matrix with a 1 in the ij-entry, zeros elsewhere, of $\mathfrak{gl}(V)$ when written with respect to this fixed basis. Consequently, $\mathrm{ad}\, x$ is diagonalisable, because we've exhibited a basis of eigenvectors.

3. Proof of Lemma 5.4(ii): want to show that $(t_\alpha, t_\alpha) \neq 0$.

Suppose that $(t_\alpha, t_\alpha) = 0$. Consider the three dimensional Lie subalgebra $S = \langle e_\alpha, t_\alpha, f_\alpha \rangle$ of L. One shows that $\mathrm{tr}_S(\mathrm{ad}_S\, x\, \mathrm{ad}_S\, y) = 0$ for all x, y in the basis $e_\alpha, t_\alpha, f_\alpha$ for S. So, by Cartan's criterion, S is a 3 dimensional soluble Lie algebra. Consequently, S' is a nilpotent Lie algebra (indeed, the derived algebra of any soluble Lie algebra over $\mathbb{C}$ is nilpotent). So, $\mathrm{ad}_L\, s$ is *nilpotent* for every $s \in S'$.

On the other hand, $\mathrm{ad}_L\, t_\alpha$ is semisimple. So, $\mathrm{ad}_L\, t_\alpha$ is actually 0 (as it is both semisimple and nilpotent). So, $[t_\alpha, L_\beta] = \beta(t_\alpha)L_\beta = 0$. This implies that α lies in the radical of the Killing form on H^*, but this is non-degenerate, a contradiction.

Jon Brundan, 21/2/97.