Commutators, matrices and an identity of Copeland

Darij Grinberg

August 24, 2019

Abstract. Given two elements a and b of a noncommutative ring, we express $(ba)^n$ as a "row vector times matrix times column vector" product, where the matrix is the n-th power of a matrix with entries $\dbinom{i}{j} \operatorname{ad}_a^{i-j}(b)$. This generalizes a formula by Tom Copeland used in the study of Pascal-style matrices.

Contents

1. Introduction

In [MO337766], Tom Copeland stated a formula for the n-th power of a differential operator. Our goal in this note is to prove a more general version of this formula, in which differential operators are replaced by arbitrary elements of a noncommutative ring.

In a nutshell, this general result (Theorem 2.7) can be stated as follows: If $n \in \mathbb{N}$ and $m \in \mathbb{N} \cup \{\infty\}$ satisfy $n < m$, and if a and b are two elements of a (noncommutative) ring $\mathbb{L}$, then

$$(ba)^n = e_0^T \left(U_b S\right)^n H_1,$$

where the column vectors e_0 and H_1 of size m are defined by

$$e_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \text{and} \qquad H_1 = \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ \vdots \\ a^{m-1} \end{pmatrix},$$

and where the $m \times m$-matrices S and U_b are defined by

$$S = ([j = i+1])_{0 \leq i < m,\ 0 \leq j < m} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \qquad \text{and}$$

$$U_b = \left(\begin{cases} \dbinom{i}{j} \operatorname{ad}_a^{i-j}(b), & \text{if } i \geq j; \\ 0, & \text{if } i < j \end{cases} \right)_{0 \leq i < m,\ 0 \leq j < m}$$

$$= \begin{pmatrix} b & 0 & 0 & \cdots & 0 \\ \operatorname{ad}_a(b) & b & 0 & \cdots & 0 \\ \operatorname{ad}_a^2(b) & 2\operatorname{ad}_a(b) & b & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \operatorname{ad}_a^{m-1}(b) & (m-1)\operatorname{ad}_a^{m-2}(b) & \dbinom{m-1}{2}\operatorname{ad}_a^{m-3}(b) & \cdots & b \end{pmatrix}$$

(using the standard Lie-algebraic notation ad_a for the operator $\mathbb{L} \to \mathbb{L}$, $c \mapsto ac - ca$). (We shall introduce all these notations in more detail below.)

Acknowledgments

DG thanks the Mathematisches Forschungsinstitut Oberwolfach for its hospitality during part of the writing process.

2. The general formula

2.1. Standing notations

Let us start by introducing notations that will remain in place for the rest of this note:

- Let $\mathbb{N}$ denote the set $\{0, 1, 2, \ldots\}$.
- "Ring" will always mean "associative ring with unity". Commutativity is not required.
- Fix a ring $\mathbb{L}$.
- For any two elements a and b of $\mathbb{L}$, we define an element $[a, b]$ of $\mathbb{L}$ by

  $$[a, b] = ab - ba.$$

  This element $[a, b]$ is called the *commutator* of a and b.

- For any $a \in \mathbb{L}$, we define a map $\operatorname{ad}_a : \mathbb{L} \to \mathbb{L}$ by
$$\left(\operatorname{ad}_a(b) = [a,b] \qquad \text{for all } b \in \mathbb{L}\right).$$
Clearly, this map ad_a is $\mathbb{Z}$-linear.

2.2. Conventions about matrices

In the following, we will use matrices. We shall use a slightly nonstandard convention for labeling the rows and the columns of our matrices: Namely, the rows and the columns of our matrices will always be indexed starting with 0. That is, a $k \times \ell$-matrix (for $k \in \mathbb{N}$ and $\ell \in \mathbb{N}$) will always have its rows numbered $0, 1, \ldots, k-1$ and its columns numbered $0, 1, \ldots, \ell - 1$. In other words, a $k \times \ell$-matrix is a family $\left(a_{i,j}\right)_{0 \leq i < k,\ 0 \leq j < \ell}$ indexed by pairs (i,j) of integers satisfying $0 \leq i < k$ and $0 \leq j < \ell$. We let $\mathbb{L}^{k \times \ell}$ denote the set of all $k \times \ell$-matrices with entries in $\mathbb{L}$.

If A is any $k \times \ell$-matrix (where k and ℓ belong to $\mathbb{N}$), and if i and j are any two integers satisfying $0 \leq i < k$ and $0 \leq j < \ell$, then we let $A_{i,j}$ denote the (i,j)-th entry of A. Thus, any $k \times \ell$-matrix A satisfies

$$A = \begin{pmatrix} A_{0,0} & A_{0,1} & \cdots & A_{0,\ell-1} \\ A_{1,0} & A_{1,1} & \cdots & A_{1,\ell-1} \\ \vdots & \vdots & \ddots & \vdots \\ A_{k-1,0} & A_{k-1,1} & \cdots & A_{k-1,\ell-1} \end{pmatrix}.$$

If $k \in \mathbb{N}$, then a *column vector of size k* means a $k \times 1$-matrix. Thus, a column vector of size k has the form $\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{pmatrix}_{0 \leq i < k,\ 0 \leq j < 1}$. Row vectors are defined similarly.

As usual, we shall equate 1×1-matrices $A \in \mathbb{L}^{1 \times 1}$ with elements of $\mathbb{L}$ (namely, by equating each 1×1-matrix $A \in \mathbb{L}^{1 \times 1}$ with its unique entry $A_{0,0}$). Thus, if v and w are any two column vectors of size k, then $w^T v \in \mathbb{L}$.

2.3. Conventions about infinite matrices

Furthermore, we shall allow our matrices to be infinite (i.e., have infinitely many rows or columns or both). This will be an optional feature of our results; we will state our claims in a way that allows the matrices to be infinite, but if the reader is only interested in finite matrices, they can ignore this possibility and skip Subsection 2.3 entirely.

First of all, let us say a few words about how we will use ∞ in this note. As usual, "∞" is just a symbol which we subject to the following rules: We have $n < \infty$ and $\infty + n = \infty - n = \infty$ for each $n \in \mathbb{N}$. Moreover, we shall use the somewhat strange convention that $\{0, 1, \ldots, \infty\}$ denotes the set $\mathbb{N}$ (so it does not contain ∞). This has the consequence that $\{0, 1, \ldots, \infty - n\} = \mathbb{N}$ for each $n \in \mathbb{N}$ (since $\infty - n = \infty$).

We will use the following kinds of infinite matrices:

- A $k \times \infty$-*matrix* (where $k \in \mathbb{N}$) has k rows (indexed by $0, 1, \ldots, k-1$) and infinitely many columns (indexed by $0, 1, 2, \ldots$). Such a matrix will usually be written as
$$\begin{pmatrix} a_{0,0} & a_{0,1} & a_{0,2} & \cdots \\ a_{1,0} & a_{1,1} & a_{1,2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ a_{k-1,0} & a_{k-1,1} & a_{k-1,2} & \cdots \end{pmatrix} = \left(a_{i,j}\right)_{0 \leq i < k,\ 0 \leq j < \infty}.$$

- A $\infty \times \ell$-*matrix* (where $\ell \in \mathbb{N}$) has infinitely many rows (indexed by $0, 1, 2, \ldots$) and ℓ columns (indexed by $0, 1, \ldots, \ell - 1$). Such a matrix will usually be written as
$$\begin{pmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,\ell-1} \\ a_{1,0} & a_{1,1} & \cdots & a_{1,\ell-1} \\ a_{2,0} & a_{2,1} & \cdots & a_{2,\ell-1} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} = \left(a_{i,j}\right)_{0 \leq i < \infty,\ 0 \leq j < \ell}.$$

- A $\infty \times \infty$-*matrix* has infinitely many rows (indexed by $0, 1, 2, \ldots$) and infinitely many columns (indexed by $0, 1, 2, \ldots$). Such a matrix will usually be written as
$$\begin{pmatrix} a_{0,0} & a_{0,1} & a_{0,2} & \cdots \\ a_{1,0} & a_{1,1} & a_{1,2} & \cdots \\ a_{2,0} & a_{2,1} & a_{2,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \left(a_{i,j}\right)_{0 \leq i < \infty,\ 0 \leq j < \infty}.$$

Matrices of these three kinds (that is, $k \times \infty$-matrices, $\infty \times \ell$-matrices and $\infty \times \infty$-matrices) will be called *infinite matrices*. In contrast, $k \times \ell$-matrices with $k, \ell \in \mathbb{N}$ will be called *finite matrices*.

We have previously introduced the notation $A_{i,j}$ for the (i, j)-th entry of A whenever A is a $k \times \ell$-matrix. The same notation will apply when A is an infinite matrix (i.e., when one or both of k and ℓ is ∞).

If u, v, w are three elements of $\mathbb{N}$, and if A is a $u \times v$-matrix, and if B is a $v \times w$-matrix, then the product AB is a $u \times w$-matrix, and its entries are given by
$$(AB)_{i,k} = \sum_{j=0}^{v-1} A_{i,j} B_{j,k} \qquad (1)$$
$$\text{for all } i \in \{0, 1, \ldots, u-1\} \text{ and } k \in \{0, 1, \ldots, w-1\}.$$

The same formula can be used to define AB when some of u, v, w are ∞ (keeping in mind that $\{0, 1, \ldots, \infty - 1\} = \mathbb{N}$), but in this case it may fail to provide a well-defined result. Indeed, if $v = \infty$, then the sum on the right hand side of (1) is infinite and thus may fail to be well-defined. Worse yet, even when products of infinite matrices are well-defined, they can fail the associativity law $(AB)C = A(BC)$. We shall not dwell on these perversions, but rather restrict ourselves to a subclass of infinite matrices which avoids them:

Definition 2.1. Let $u, v \in \mathbb{N} \cup \{\infty\}$. Let A be a $u \times v$-matrix. Let $k \in \mathbb{Z}$. We say that the matrix A is *k-lower-triangular* if and only if we have

$$\left(A_{i,j} = 0 \qquad \text{for all } (i,j) \text{ satisfying } i < j + k\right).$$

Definition 2.2. A matrix A is said to be *quasi-lower-triangular* if and only if there exists a $k \in \mathbb{Z}$ such that A is k-lower-triangular.

Note that we did not require our matrix A to be square in these two definitions. Unlike the standard kind of triangularity, our concept of quasi-triangularity is meant to be a tameness condition, meant to guarantee the well-definedness of an infinite sum; in particular, all finite matrices are quasi-lower-triangular. Better yet, the following holds:[1]

Proposition 2.3. Let $k \in \mathbb{N} \cup \{\infty\}$ and $\ell \in \mathbb{N}$. Then, any $k \times \ell$-matrix is quasi-lower-triangular. More concretely: Any $k \times \ell$-matrix is $(\ell - 1)$-lower-triangular.

Proposition 2.4. Let A be a matrix (finite or infinite) such that all but finitely many entries of A are 0. Then, A is quasi-lower-triangular.

Quasi-lower-triangular matrices can be multiplied, as the following proposition shows:

Proposition 2.5. Let $u, v, w \in \mathbb{N} \cup \{\infty\}$. Let A be a quasi-lower-triangular $u \times v$-matrix, and let B be a quasi-lower-triangular $v \times w$-matrix. Then, the product AB is well-defined (i.e. the infinite sum on the right hand side of (1) is well-defined even if $v = \infty$) and is a quasi-lower-triangular $u \times w$-matrix.

More concretely: If $k, \ell \in \mathbb{Z}$ are such that A is k-lower-triangular and B is ℓ-lower-triangular, then AB is $(k + \ell)$-lower-triangular.

Finally, multiplication of quasi-lower-triangular matrices is associative:

Proposition 2.6. Let $u, v, w, x \in \mathbb{N} \cup \{\infty\}$. Let A be a quasi-lower-triangular $u \times v$-matrix; let B be a quasi-lower-triangular $v \times w$-matrix; let C be a quasi-lower-triangular $w \times x$-matrix. Then, $(AB)C = A(BC)$.

This proposition entails that we can calculate with quasi-lower-triangular matrices just as we can calculate with finite matrices. In particular, the quasi-lower-triangular $\infty \times \infty$-matrices form a ring. Thus, a quasi-lower-triangular $\infty \times \infty$-matrix has a well-defined n-th power for each $n \in \mathbb{N}$.

[1] The proofs of all propositions stated in Subsection 2.3 are left to the reader as easy exercises.

2.4. The matrices S and U_b and the vectors H_c and e_j

Let us now introduce several more players into the drama.

2.4.1. Iverson brackets (truth values)

We shall use the *Iverson bracket notation*: If $\mathcal{A}$ is any logical statement, then $[\mathcal{A}]$ will denote the integer $\begin{cases} 1, & \text{if } \mathcal{A} \text{ is true;} \\ 0, & \text{if } \mathcal{A} \text{ is false} \end{cases} \in \{0,1\}$. This integer $[\mathcal{A}]$ is called the *truth value* of $\mathcal{A}$.

2.4.2. m and a

We now return to our ring $\mathbb{L}$.

For the rest of this note, we fix an $m \in \mathbb{N} \cup \{\infty\}$ and an element $a \in \mathbb{L}$.

2.4.3. The matrix S

We define an $m \times m$-matrix $S \in \mathbb{L}^{m\times m}$ by

$$S = ([j = i+1])_{0\leq i<m,\ 0\leq j<m}\,. \tag{2}$$

This matrix S looks as follows:

- If $m \in \mathbb{N}$, then
$$S = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$
- If $m = \infty$, then
$$S = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The matrix S (or, rather, the $\mathbb{L}$-linear map from $\mathbb{L}^m$ to $\mathbb{L}^m$ it represents[2]) is often called the *shift operator*. Note that the matrix S is quasi-lower-triangular[3] (and, in

[2] When $m = \infty$, you can read $\mathbb{L}^m$ both as the direct sum $\bigoplus\limits_{i\in\mathbb{N}} \mathbb{L}$ and as the direct product $\prod\limits_{i\in\mathbb{N}} \mathbb{L}$. These are two different options, but either has an $\mathbb{L}$-linear map represented by the matrix S.

[3] See Subsection 2.3 for the meaning of this word (and ignore it if you don't care about the case of $m = \infty$).

fact, (-1)-lower-triangular[4]), but of course not lower-triangular (unless $\mathbb{L} = 0$ or $m \leq 1$).

2.4.4. The matrix U_b

If n is a nonnegative integer, T is a set and $f : T \to T$ is any map, then f^n will mean the composition $\underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times}}$; this is again a map from T to T.

For any $b \in \mathbb{L}$, we define an $m \times m$-matrix $U_b \in \mathbb{L}^{m \times m}$ by

$$U_b = \left(\begin{cases} \dbinom{i}{j} \operatorname{ad}_a^{i-j}(b), & \text{if } i \geq j; \\ 0, & \text{if } i < j \end{cases} \right)_{0 \leq i < m,\ 0 \leq j < m} . \tag{3}$$

(Here, of course, ad_a^n means $(\operatorname{ad}_a)^n$ whenever $n \in \mathbb{N}$.)

This matrix U_b looks as follows:

- If $b \in \mathbb{L}$ and $m \in \mathbb{N}$, then

$$U_b = \begin{pmatrix} b & 0 & 0 & \cdots & 0 \\ \operatorname{ad}_a(b) & b & 0 & \cdots & 0 \\ \operatorname{ad}_a^2(b) & 2\operatorname{ad}_a(b) & b & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \operatorname{ad}_a^{m-1}(b) & (m-1)\operatorname{ad}_a^{m-2}(b) & \dbinom{m-1}{2}\operatorname{ad}_a^{m-3}(b) & \cdots & b \end{pmatrix}.$$

- If $b \in \mathbb{L}$ and $m = \infty$, then

$$U_b = \begin{pmatrix} b & 0 & 0 & 0 & \cdots \\ \operatorname{ad}_a(b) & b & 0 & 0 & \cdots \\ \operatorname{ad}_a^2(b) & 2\operatorname{ad}_a(b) & b & 0 & \cdots \\ \operatorname{ad}_a^3(b) & 3\operatorname{ad}_a^2(b) & 3\operatorname{ad}_a(b) & b & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Note that the matrix U_b is always lower-triangular and thus quasi-lower-triangular[5].

[4] See Subsection 2.3 for the meaning of this word (and ignore it if you don't care about the case of $m = \infty$).

[5] See Subsection 2.3 for the meaning of this word (and ignore it if you don't care about the case of $m = \infty$).

2.4.5. The column vector H_c

Furthermore, for each $c \in \mathbb{L}$, we define an $m \times 1$-matrix $H_c \in \mathbb{L}^{m \times 1}$ by

$$H_c = \left(a^i c\right)_{0 \leq i < m,\ 0 \leq j < 1}. \tag{4}$$

Thus, H_c is an $m \times 1$-matrix, i.e., a column vector of size m. It looks as follows:

- If $c \in \mathbb{L}$ and $m \in \mathbb{N}$, then
$$H_c = \begin{pmatrix} a^0 c \\ a^1 c \\ \vdots \\ a^{m-1} c \end{pmatrix}.$$
- If $c \in \mathbb{L}$ and $m = \infty$, then
$$H_c = \begin{pmatrix} a^0 c \\ a^1 c \\ a^2 c \\ \vdots \end{pmatrix}.$$

Clearly, the matrix H_c is quasi-lower-triangular[6], since it has only one column.

2.4.6. The column vector e_j

For each integer j with $0 \leq j < m$, we let $e_j \in \mathbb{L}^{m \times 1}$ be the $m \times 1$-matrix defined by

$$e_j = ([p = j])_{0 \leq p < m,\ 0 \leq q < 1}. \tag{5}$$

In other words, e_j is the column vector (of size m) whose j-th entry is 1 and whose all other entries are 0. This column vector e_j is commonly known as the j-th *standard basis vector* of $\mathbb{L}^{m \times 1}$.

Thus, in particular, e_0 is a column vector with a 1 in its topmost position and 0's everywhere else. It looks as follows:

- If $m \in \mathbb{N}$, then
$$e_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

[6] See Subsection 2.3 for the meaning of this word (and ignore it if you don't care about the case of $m = \infty$).

- If $m = \infty$, then
$$e_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}.$$

Thus, e_0^T is a row vector with a 1 in its leftmost position and 0's everywhere else. This shows that the matrix e_0^T is quasi-lower-triangular[7].

2.5. The general formula

We are now ready to state our main claim:

Theorem 2.7. Let $n \in \mathbb{N}$ be such that $n < m$. Let $b \in \mathbb{L}$. Then,
$$(ba)^n = e_0^T (U_b S)^n H_1.$$

(The right hand side of this equality is a 1×1-matrix, while the left hand side is an element of $\mathbb{L}$. The equality thus makes sense because we are equating 1×1-matrices with elements of $\mathbb{L}$.)

Example 2.8. Let us set $m = 3$ and $n = 2$ in Theorem 2.7. Then, Theorem 2.7 claims that $(ba)^2 = e_0^T (U_b S)^2 H_1$. Let us check this: We have
$$U_b = \begin{pmatrix} b & 0 & 0 \\ \operatorname{ad}_a(b) & b & 0 \\ \operatorname{ad}_a^2(b) & 2\operatorname{ad}_a(b) & b \end{pmatrix} \qquad \text{and} \qquad S = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$
so that
$$U_b S = \begin{pmatrix} b & 0 & 0 \\ \operatorname{ad}_a(b) & b & 0 \\ \operatorname{ad}_a^2(b) & 2\operatorname{ad}_a(b) & b \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & b & 0 \\ 0 & \operatorname{ad}_a(b) & b \\ 0 & \operatorname{ad}_a^2(b) & 2\operatorname{ad}_a(b) \end{pmatrix}$$
and therefore
$$\begin{aligned}(U_b S)^2 &= \begin{pmatrix} 0 & b & 0 \\ 0 & \operatorname{ad}_a(b) & b \\ 0 & \operatorname{ad}_a^2(b) & 2\operatorname{ad}_a(b) \end{pmatrix}^2 \\ &= \begin{pmatrix} 0 & b\operatorname{ad}_a(b) & b^2 \\ 0 & (\operatorname{ad}_a(b))^2 + b\operatorname{ad}_a^2(b) & 3b\operatorname{ad}_a(b) \\ 0 & 3\operatorname{ad}_a(b)\operatorname{ad}_a^2(b) & 4(\operatorname{ad}_a(b))^2 + b\operatorname{ad}_a^2(b) \end{pmatrix}.\end{aligned}$$

[7] See Subsection 2.3 for the meaning of this word (and ignore it if you don't care about the case of $m = \infty$).

Multiplying $e_0^T = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$ by this equality, we find

$$\begin{aligned} e_0^T (U_b S)^2 &= \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & b \operatorname{ad}_a(b) & b^2 \\ 0 & (\operatorname{ad}_a(b))^2 + b \operatorname{ad}_a^2(b) & 3b \operatorname{ad}_a(b) \\ 0 & 3 \operatorname{ad}_a(b) \operatorname{ad}_a^2(b) & 4(\operatorname{ad}_a(b))^2 + b \operatorname{ad}_a^2(b) \end{pmatrix} \\ &= \begin{pmatrix} 0 & b \operatorname{ad}_a(b) & b^2 \end{pmatrix}. \end{aligned}$$

Multiplying this equality by $H_1 = \begin{pmatrix} a^0 1 \\ a^1 1 \\ a^2 1 \end{pmatrix} = \begin{pmatrix} a^0 \\ a^1 \\ a^2 \end{pmatrix}$, we obtain

$$\begin{aligned} e_0^T (U_b S)^2 H_1 &= \begin{pmatrix} 0 & b \operatorname{ad}_a(b) & b^2 \end{pmatrix} \begin{pmatrix} a^0 \\ a^1 \\ a^2 \end{pmatrix} = 0a^0 + b \operatorname{ad}_a(b) a^1 + b^2 a^2 \\ &= b \underbrace{\operatorname{ad}_a(b)}_{\substack{=[a,b] \\ \text{(by the definition of } \operatorname{ad}_a\text{)}}} a + b^2 a^2 = b \underbrace{[a,b]}_{=ab-ba} a + b^2 a^2 \\ &= b(ab - ba)a + b^2 a^2 = baba - bbaa + bbaa = baba = (ba)^2. \end{aligned}$$

This confirms the claim that $(ba)^2 = e_0^T (U_b S)^2 H_1$.

3. The proof

3.1. The idea

Proving Theorem 2.7 is not hard, but it will take us some preparation due to the bookkeeping required. The main idea manifests itself in its cleanest form when $m = \infty$; indeed, it is not hard to prove the following two facts:[8]

Proposition 3.1. Assume that $m = \infty$. Let $c \in \mathbb{L}$. Then, $SH_c = H_{ac}$.

Proposition 3.2. Let $b \in \mathbb{L}$ and $c \in \mathbb{L}$. Then, $U_b H_c = H_{bc}$.

If $m = \infty$, then we can use Proposition 3.1 and Proposition 3.2 to conclude that $(U_b S) H_c = H_{bac}$ for each $b \in \mathbb{L}$ and $c \in \mathbb{L}$. Thus, by induction, we can conclude that $(U_b S)^n H_c = H_{(ba)^n c}$ for each $n \in \mathbb{N}$, $b \in \mathbb{L}$ and $c \in \mathbb{L}$ (as long as $m = \infty$). Applying this to $c = 1$ and multiplying the resulting equality by e_0^T on both sides, we then obtain $e_0^T (U_b S)^n H_1 = e_0^T H_{(ba)^n 1} = (ba)^n$ (the last equality sign is easy). This proves Theorem 2.7 in the case when $m = \infty$.

[8] We shall prove these two facts later.

Unfortunately, this argument breaks down if $m \in \mathbb{N}$. In fact, Proposition 3.1 is true only for $m = \infty$; otherwise, the vectors SH_c and H_{ac} differ in their last entry. This "corruption" then spreads further to earlier and earlier entries as we inductively multiply by U_b and by S. What saves us is that it only spreads one entry at a time when we multiply by S, and does not spread at all when we multiply by U_b; thus it does not reach the first (i.e., 0-th) entry as long as we multiply by U_bS only n times. But this needs to be formalized and proved. This is what we shall be doing further below.

3.2. A lemma about ad_a

Before we come to this, however, we need a basic lemma about commutators:

Lemma 3.3. Let $b \in \mathbb{L}$ and $i \in \mathbb{N}$. Then,

$$a^i b = \sum_{j=0}^{i} \binom{i}{j} \operatorname{ad}_a^{i-j}(b) \cdot a^j.$$

It is not hard to prove Lemma 3.3 by induction on i. However, there is a slicker proof. It relies on the following well-known fact:

Proposition 3.4. Let $\mathbb{A}$ be a ring. Let x and y be two elements of $\mathbb{A}$ such that $xy = yx$. Then,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \qquad \text{for every } n \in \mathbb{N}.$$

Proposition 3.4 is a straightforward generalization of the binomial formula to two commuting elements of an arbitrary ring.

Proof of Lemma 3.3. Let $\operatorname{End}\mathbb{L}$ denote the endomorphism ring of the $\mathbb{Z}$-module $\mathbb{L}$. Thus, the elements of $\operatorname{End}\mathbb{L}$ are the $\mathbb{Z}$-linear maps from $\mathbb{L}$ to $\mathbb{L}$.

Define the map $L_a : \mathbb{L} \to \mathbb{L}$ by

$$(L_a(c) = ac \qquad \text{for all } c \in \mathbb{L}).$$

Clearly, this map L_a is $\mathbb{Z}$-linear; thus, it belongs to $\operatorname{End}\mathbb{L}$.

Define the map $R_a : \mathbb{L} \to \mathbb{L}$ by

$$(R_a(c) = ca \qquad \text{for all } c \in \mathbb{L}).$$

Clearly, this map R_a is $\mathbb{Z}$-linear; thus, it belongs to $\operatorname{End}\mathbb{L}$.

We have $\operatorname{ad}_a = L_a - R_a$ [9]. Hence, ad_a belongs to $\operatorname{End}\mathbb{L}$ (since L_a and R_a belong to $\operatorname{End}\mathbb{L}$). Also, $R_a + \operatorname{ad}_a = L_a$ (since $\operatorname{ad}_a = L_a - R_a$).

Furthermore, the elements L_a and R_a of $\operatorname{End}\mathbb{L}$ satisfy $R_a \circ L_a = L_a \circ R_a$ [10]. But $\operatorname{End}\mathbb{L}$ is a ring with multiplication $\circ$; thus, in particular, the operation $\circ$ is distributive (over $+$) on $\operatorname{End}\mathbb{L}$. Since L_a, R_a and ad_a belong to $\operatorname{End}\mathbb{L}$, we thus have

$$\begin{aligned} R_a \circ \underbrace{\operatorname{ad}_a}_{=L_a - R_a} &= R_a \circ (L_a - R_a) = \underbrace{R_a \circ L_a}_{=L_a \circ R_a} - R_a \circ R_a \\ &= L_a \circ R_a - R_a \circ R_a = \underbrace{(L_a - R_a)}_{=\operatorname{ad}_a} \circ R_a = \operatorname{ad}_a \circ R_a. \end{aligned}$$

Hence, Proposition 3.4 (applied to $\mathbb{A} = \operatorname{End}\mathbb{L}$, $x = R_a$, $y = \operatorname{ad}_a$ and $n = i$) yields

$$(R_a + \operatorname{ad}_a)^i = \sum_{k=0}^{i} \binom{i}{k} R_a^k \circ \operatorname{ad}_a^{i-k} = \sum_{j=0}^{i} \binom{i}{j} R_a^j \circ \operatorname{ad}_a^{i-j}$$

(here, we have renamed the index k as j in the sum). In view of $R_a + \operatorname{ad}_a = L_a$, this rewrites as

$$L_a^i = \sum_{j=0}^{i} \binom{i}{j} R_a^j \circ \operatorname{ad}_a^{i-j}. \tag{6}$$

But each $k \in \mathbb{N}$ satisfies

$$L_a^k (c) = a^k c \qquad \text{for each } c \in \mathbb{L}. \tag{7}$$

[9] *Proof.* Let $c \in \mathbb{L}$. Then, $L_a(c) = ac$ (by the definition of L_a) and $R_a(c) = ca$ (by the definition of R_a). Hence,

$$(L_a - R_a)(c) = \underbrace{L_a(c)}_{=ac} - \underbrace{R_a(c)}_{=ca} = ac - ca.$$

Comparing this with

$$\begin{aligned} \operatorname{ad}_a(c) &= [a, c] && \text{(by the definition of } \operatorname{ad}_a) \\ &= ac - ca && \text{(by the definition of } [a, c]), \end{aligned}$$

we obtain $\operatorname{ad}_a(c) = (L_a - R_a)(c)$.

Now, forget that we fixed c. We thus have shown that $\operatorname{ad}_a(c) = (L_a - R_a)(c)$ for each $c \in \mathbb{L}$. In other words, $\operatorname{ad}_a = L_a - R_a$. Qed.

[10] *Proof.* Let $c \in \mathbb{L}$. The definition of L_a yields $L_a(c) = ac$ and $L_a(R_a(c)) = a \cdot R_a(c)$. The definition of R_a yields $R_a(c) = ca$ and $R_a(L_a(c)) = L_a(c) \cdot a$. Now, comparing

$$(L_a \circ R_a)(c) = L_a(R_a(c)) = a \cdot \underbrace{R_a(c)}_{=ca} = a \cdot ca = aca$$

with

$$(R_a \circ L_a)(c) = R_a(L_a(c)) = \underbrace{L_a(c)}_{=ac} \cdot a = ac \cdot a = aca,$$

we obtain $(R_a \circ L_a)(c) = (L_a \circ R_a)(c)$.

Forget that we fixed c. We thus have proven that $(R_a \circ L_a)(c) = (L_a \circ R_a)(c)$ for each $c \in \mathbb{L}$. In other words, $R_a \circ L_a = L_a \circ R_a$.

[*Proof of (7):* It is straightforward to prove (7) by induction on k.] Furthermore, each $k \in \mathbb{N}$ satisfies

$$R_a^k(c) = ca^k \qquad \text{for each } c \in \mathbb{L}. \tag{8}$$

[*Proof of (8):* It is straightforward to prove (8) by induction on k.] Now, applying both sides of the equality (6) to b, we obtain

$$L_a^i(b) = \left(\sum_{j=0}^{i} \binom{i}{j} R_a^j \circ \operatorname{ad}_a^{i-j}\right)(b) = \sum_{j=0}^{i} \binom{i}{j} \underbrace{\left(R_a^j \circ \operatorname{ad}_a^{i-j}\right)(b)}_{\substack{=R_a^j\left(\operatorname{ad}_a^{i-j}(b)\right) \\ =\operatorname{ad}_a^{i-j}(b)\cdot a^j \\ \text{(by (8), applied} \\ \text{to } k=j \text{ and } c=\operatorname{ad}_a^{i-j}(b))}}$$

$$= \sum_{j=0}^{i} \binom{i}{j} \operatorname{ad}_a^{i-j}(b) \cdot a^j.$$

Comparing this with

$$L_a^i(b) = a^i b \qquad \text{(by (7), applied to } k = i \text{ and } c = b),$$

we obtain

$$a^i b = \sum_{j=0}^{i} \binom{i}{j} \operatorname{ad}_a^{i-j}(b) \cdot a^j.$$

This proves Lemma 3.3. □

3.3. Formulas for $e_i^T A$

We next recall a simple property of the vectors e_i:

Lemma 3.5. Let $\ell \in \mathbb{N} \cup \{\infty\}$ and $i \in \mathbb{N}$ be such that $0 \leq i < m$. Let A be an $m \times \ell$-matrix. Then,

$$e_i^T A = (\text{the } i\text{-th row of } A).$$

Note that the product $e_i^T A$ on the left hand side of Lemma 3.5 is always well-defined, even when ℓ and m are ∞. (This stems from the fact that the row vector e_i^T has only one nonzero entry.)

Lemma 3.5 says that the i-th row of A can be extracted by multiplying A from the left by the row vector $e_i^T = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$ (here, the 1 is at the i-th position). This is a known fact from linear algebra and is easy to prove.

The next lemma is a slight restatement of Lemma 3.5 in the case when $\ell = m$:

Lemma 3.6. Let $i \in \mathbb{N}$ be such that $0 \leq i < m$. Let A be an $m \times m$-matrix. Then,

$$e_i^T A = \sum_{j=0}^{m-1} A_{i,j} e_j^T.$$

Proof of Lemma 3.6. For each $j \in \{0, 1, \ldots, m-1\}$, we have $e_j = ([p = j])_{0 \leq p < m,\ 0 \leq q < 1}$ (by (5)) and thus

$$e_j^T = ([q = j])_{0 \leq p < 1,\ 0 \leq q < m} \qquad \text{(by the definition of the transpose of a matrix)}.$$

Hence,

$$\begin{aligned} \sum_{j=0}^{m-1} A_{i,j} \underbrace{e_j^T}_{=([q=j])_{0 \leq p < 1,\ 0 \leq q < m}} &= \sum_{j=0}^{m-1} A_{i,j} ([q = j])_{0 \leq p < 1,\ 0 \leq q < m} \\ &= \left(\sum_{j=0}^{m-1} A_{i,j} [q = j] \right)_{0 \leq p < 1,\ 0 \leq q < m}. \end{aligned} \tag{9}$$

But for each $q \in \{0, 1, \ldots, m-1\}$, we have

$$\begin{aligned} \sum_{j=0}^{m-1} A_{i,j} [q = j] &= A_{i,q} \underbrace{[q = q]}_{\substack{=1 \\ \text{(since } q=q\text{)}}} + \sum_{\substack{j \in \{0,1,\ldots,m-1\}; \\ j \neq q}} A_{i,j} \underbrace{[q = j]}_{\substack{=0 \\ \text{(because } j \neq q\text{)}}} \\ &\qquad \left(\begin{array}{c} \text{here, we have split off the addend for } j = q \\ \text{from the sum (since } q \in \{0, 1, \ldots, m-1\}\text{)} \end{array} \right) \\ &= A_{i,q} + \underbrace{\sum_{\substack{j \in \{0,1,\ldots,m-1\}; \\ j \neq q}} A_{i,j} 0}_{=0} = A_{i,q}. \end{aligned}$$

Hence,

$$\left(\sum_{j=0}^{m-1} A_{i,j} [q = j] \right)_{0 \leq p < 1,\ 0 \leq q < m} = \left(A_{i,q} \right)_{0 \leq p < 1,\ 0 \leq q < m} = (\text{the } i\text{-th row of } A)$$

(since $A = \left(A_{i,j} \right)_{0 \leq i < m,\ 0 \leq j < m}$). Hence, (9) becomes

$$\sum_{j=0}^{m-1} A_{i,j} e_j^T = \left(\sum_{j=0}^{m-1} A_{i,j} [q = j] \right)_{0 \leq p < 1,\ 0 \leq q < m} = (\text{the } i\text{-th row of } A) = e_i^T A$$

(since Lemma 3.5 yields $e_i^T A = (\text{the } i\text{-th row of } A)$). This proves Lemma 3.6. □

3.4. Proving $e_u^T S H_c = e_u^T H_{ac}$ for $u+1 < m$

We can now prove a generalization of Proposition 3.1 to the case of arbitrary m:

> **Proposition 3.7.** Let $u \in \mathbb{N}$ be such that $u+1 < m$. Then:
> **(a)** We have $e_u^T S = e_{u+1}^T$.
> **(b)** Let $c \in \mathbb{L}$. Then, $e_u^T S H_c = e_u^T H_{ac}$.

Proof of Proposition 3.7. **(a)** Lemma 3.5 (applied to $\ell = m$, $A = S$ and $i = u$) yields

$$e_u^T S = (\text{the } u\text{-th row of } S) = ([q = u+1])_{0 \le p < 1,\ 0 \le q < m} \qquad (10)$$

(by (2)). But (5) (applied to $j = u+1$) yields

$$e_{u+1} = ([p = u+1])_{0 \le p < m,\ 0 \le q < 1}.$$

Thus, by the definition of the transpose of a matrix, we obtain

$$e_{u+1}^T = ([q = u+1])_{0 \le p < 1,\ 0 \le q < m}.$$

Comparing this with (10), we obtain $e_u^T S = e_{u+1}^T$. This proves Proposition 3.7 **(a)**.

(b) Lemma 3.5 (applied to $\ell = 1$, $A = H_{ac}$ and $i = u$) yields

$$\begin{aligned} e_u^T H_{ac} &= (\text{the } u\text{-th row of } H_{ac}) \\ &= (\text{the } u\text{-th entry of } H_{ac}) \qquad (\text{since } H_{ac} \text{ is a column vector}) \\ &= \underbrace{a^u a}_{=a^{u+1}} c \qquad \left(\text{since (4) yields } H_{ac} = \left(a^i ac\right)_{0 \le i < m,\ 0 \le j < 1}\right) \\ &= a^{u+1} c. \end{aligned}$$

Comparing this with

$$\begin{aligned} &\underbrace{e_u^T S}_{\substack{=e_{u+1}^T \\ \text{(by Proposition 3.7 (a))}}} H_c \\ &= e_{u+1}^T H_c = (\text{the } (u+1)\text{-th row of } H_c) \\ &\qquad (\text{by Lemma 3.5, applied to } \ell = 1,\ A = H_c \text{ and } i = u+1) \\ &= (\text{the } (u+1)\text{-th entry of } H_c) \qquad (\text{since } H_c \text{ is a column vector}) \\ &= a^{u+1} c \qquad \left(\text{since (4) yields } H_c = \left(a^i c\right)_{0 \le i < m,\ 0 \le j < 1}\right), \end{aligned}$$

we obtain $e_u^T S H_c = e_u^T H_{ac}$. This proves Proposition 3.7 **(b)**. □

It is now easy to derive Proposition 3.1 from Proposition 3.7 **(b)**:

Proof of Proposition 3.1 (sketched). We have $m = \infty$; thus, every $u \in \mathbb{N}$ satisfies $u + 1 < m$. Hence, Proposition 3.7 **(b)** yields that $e_u^T S H_c = e_u^T H_{ac}$ for every $u \in \mathbb{N}$. From this, it is easy to conclude that $S H_c = H_{ac}$ (using Lemma 3.5). We leave the details to the reader, since we will not use Proposition 3.1. □

3.5. Proving $U_b H_c = H_{bc}$

Next, we shall prove Proposition 3.2. For convenience, let us recall its statement:

Proposition 3.8. Let $b \in \mathbb{L}$ and $c \in \mathbb{L}$. Then, $U_b H_c = H_{bc}$.

Proof of Proposition 3.8. Let $u \in \{0, 1, \ldots, m-1\}$. Hence, $0 \leq u \leq m-1$. (Keep in mind that $\{0, 1, \ldots, \infty - 1\} = \mathbb{N}$, so u cannot be ∞ even when $m = \infty$.)

From (3), we see that

$$\left(U_b\right)_{i,j} = \begin{cases} \dbinom{i}{j} \operatorname{ad}_a^{i-j}(b), & \text{if } i \geq j; \\ 0, & \text{if } i < j \end{cases} \tag{11}$$

for each $i \in \{0, 1, \ldots, m-1\}$ and $j \in \{0, 1, \ldots, m-1\}$.

From (4), we obtain

$$\left(H_c\right)_{i,0} = a^i c \tag{12}$$

for each $i \in \{0, 1, \ldots, m-1\}$. The same argument (applied to bc instead of c) yields

$$\left(H_{bc}\right)_{i,0} = a^i bc \tag{13}$$

for each $i \in \{0, 1, \ldots, m-1\}$.

Now, (1) (applied to m, m, 1, U_b, H_c, u and 0 instead of u, v, w, A, B, i and k) yields

$$\begin{aligned}
\left(U_b H_c\right)_{u,0} &= \sum_{j=0}^{m-1} \underbrace{\left(U_b\right)_{u,j}}_{\substack{= \begin{cases} \binom{u}{j} \operatorname{ad}_a^{u-j}(b), & \text{if } u \geq j; \\ 0, & \text{if } u < j \end{cases} \\ \text{(by (11), applied to } i = u)}} \underbrace{\left(H_c\right)_{j,0}}_{\substack{= a^j c \\ \text{(by (12), applied to } i = j)}} \\
&= \sum_{j=0}^{m-1} \begin{cases} \dbinom{u}{j} \operatorname{ad}_a^{u-j}(b), & \text{if } u \geq j; \\ 0, & \text{if } u < j \end{cases} \cdot a^j c \\
&= \sum_{j=0}^{u} \underbrace{\begin{cases} \dbinom{u}{j} \operatorname{ad}_a^{u-j}(b), & \text{if } u \geq j; \\ 0, & \text{if } u < j \end{cases}}_{\substack{= \binom{u}{j} \operatorname{ad}_a^{u-j}(b) \\ \text{(since } u \geq j \text{ (because } j \leq u))}} \cdot a^j c + \sum_{j=u+1}^{m-1} \underbrace{\begin{cases} \dbinom{u}{j} \operatorname{ad}_a^{u-j}(b), & \text{if } u \geq j; \\ 0, & \text{if } u < j \end{cases}}_{\substack{= 0 \\ \text{(since } u < j \text{ (because } j \geq u+1 > u))}} \cdot a^j c \\
&\qquad \left(\text{here, we have split the sum at } j = u \text{, since } 0 \leq u \leq m-1\right) \\
&= \sum_{j=0}^{u} \binom{u}{j} \operatorname{ad}_a^{u-j}(b) \cdot a^j c + \underbrace{\sum_{j=u+1}^{m-1} 0 \cdot a^j c}_{=0} = \sum_{j=0}^{u} \binom{u}{j} \operatorname{ad}_a^{u-j}(b) \cdot a^j c.
\end{aligned}$$

Comparing this with

$$\begin{aligned}(H_{bc})_{u,0} &= \underbrace{a^u b}_{\substack{=\sum\limits_{j=0}^{u}\binom{u}{j}\operatorname{ad}_a^{u-j}(b)\cdot a^j \\ \text{(by Lemma 3.3, applied to } i=u\text{)}}} c \qquad \text{(by (13), applied to } i=u\text{)} \\ &= \left(\sum_{j=0}^{u}\binom{u}{j}\operatorname{ad}_a^{u-j}(b)\cdot a^j\right)c = \sum_{j=0}^{u}\binom{u}{j}\operatorname{ad}_a^{u-j}(b)\cdot a^j c,\end{aligned}$$

we obtain $(U_b H_c)_{u,0} = (H_{bc})_{u,0}$.

Now, recall that $U_b H_c$ is a column vector. Hence,

$$(\text{the } u\text{-th entry of } U_b H_c) = (U_b H_c)_{u,0} = (H_{bc})_{u,0}\,. \tag{14}$$

But H_{bc} is also a column vector. Thus,

$$(\text{the } u\text{-th entry of } H_{bc}) = (H_{bc})_{u,0}\,.$$

Comparing this with (14), we obtain

$$(\text{the } u\text{-th entry of } U_b H_c) = (\text{the } u\text{-th entry of } H_{bc})\,.$$

Now, forget that we fixed u. We thus have shown that (the u-th entry of $U_b H_c$) $=$ (the u-th entry of H_{bc}) for each $u \in \{0, 1, \ldots, m-1\}$. In other words, each entry of $U_b H_c$ equals the corresponding entry of H_{bc}. Thus, the two column vectors $U_b H_c$ and H_{bc} are identical. In other words, $U_b H_c = H_{bc}$. This proves Proposition 3.8. □

3.6. The $\underset{k}{\equiv}$ relations

Now, we introduce a notation for saying that two $m \times \ell$-matrices are equal in their first $m - k + 1$ rows:

> **Definition 3.9.** Let $\ell \in \mathbb{N} \cup \{\infty\}$. Let $A \in \mathbb{L}^{m\times\ell}$ and $B \in \mathbb{L}^{m\times\ell}$ be two $m \times \ell$-matrices. Let k be a positive integer. We shall say that $A \underset{k}{\equiv} B$ if and only if we have
>
> $$\left(e_u^T A = e_u^T B \qquad \text{for all } u \in \{0, 1, \ldots, m-k\}\right).$$

(Recall again that $\{0, 1, \ldots, \infty\}$ means $\mathbb{N}$; thus, "$u \in \{0, 1, \ldots, m-k\}$" means "$u \in \mathbb{N}$" in the case when $m = \infty$. Note that $\{0, 1, \ldots, g\}$ means the empty set $\varnothing$ when $g < 0$.)

Note that the condition "$e_u^T A = e_u^T B$" in Definition 3.9 can be restated as "the u-th row of A equals the u-th row of B", because of Lemma 3.5. But we will find it easier to use it in the form "$e_u^T A = e_u^T B$".

The following lemma is easy:

Lemma 3.10. Let $\ell \in \mathbb{N} \cup \{\infty\}$. Let $A \in \mathbb{L}^{m\times \ell}$ and $B \in \mathbb{L}^{m\times \ell}$ be two $m \times \ell$-matrices. Let k be a positive integer such that $A \underset{k}{\equiv} B$. Let $b \in \mathbb{L}$. Then, $U_b A \underset{k}{\equiv} U_b B$.

All that is needed of the matrix U_b for Lemma 3.10 to hold is that U_b is lower-triangular; we stated it for U_b just for convenience reasons.

Proof of Lemma 3.10. We have $A \underset{k}{\equiv} B$. In other words, we have

$$\left(e_u^T A = e_u^T B \qquad \text{for all } u \in \{0, 1, \ldots, m-k\}\right) \tag{15}$$

(by the definition of "$A \underset{k}{\equiv} B$").

Let $u \in \{0, 1, \ldots, m-k\}$. Thus, $0 \leq u \leq m-k$. Also, $u \in \{0, 1, \ldots, m-k\} \subseteq \{0, 1, \ldots, m-1\}$ (since $m - \underbrace{k}_{\geq 1} \leq m-1$). Hence, $0 \leq u \leq m-1 < m$. For each $j \in \{0, 1, \ldots, u\}$, we have $j \in \{0, 1, \ldots, u\} \subseteq \{0, 1, \ldots, m-k\}$ (since $u \leq m-k$) and thus

$$e_j^T A = e_j^T B \tag{16}$$

(by (15), applied to j instead of u).

From (3), we see that

$$(U_b)_{i,j} = \begin{cases} \dbinom{i}{j} \operatorname{ad}_a^{i-j}(b), & \text{if } i \geq j; \\ 0, & \text{if } i < j \end{cases} \tag{17}$$

for each $i \in \{0, 1, \ldots, m-1\}$ and $j \in \{0, 1, \ldots, m-1\}$.

For each $j \in \{u+1, u+2, \ldots, m-1\}$, we have $j \geq u+1$ and

$$\begin{aligned} (U_b)_{u,j} &= \begin{cases} \dbinom{u}{j} \operatorname{ad}_a^{u-j}(b), & \text{if } u \geq j; \\ 0, & \text{if } u < j \end{cases} \qquad \text{(by (17), applied to } i = u\text{)} \\ &= 0 \qquad (\text{since } u < j \text{ (because } j \geq u+1 > u)). \end{aligned} \tag{18}$$

Now, Lemma 3.6 (applied to $i = u$ and $A = U_b$) yields

$$\begin{aligned} e_u^T U_b &= \sum_{j=0}^{m-1} (U_b)_{u,j} e_j^T = \sum_{j=0}^{u} (U_b)_{u,j} e_j^T + \sum_{j=u+1}^{m-1} \underbrace{(U_b)_{u,j}}_{\substack{=0 \\ \text{(by (18))}}} e_j^T \\ &\qquad \left(\text{here, we have split the sum at } j = u, \text{ since } 0 \leq u \leq m-1\right) \\ &= \sum_{j=0}^{u} (U_b)_{u,j} e_j^T + \underbrace{\sum_{j=u+1}^{m-1} 0 e_j^T}_{=0} = \sum_{j=0}^{u} (U_b)_{u,j} e_j^T. \end{aligned}$$

Hence,

$$\underbrace{e_u^T U_b}_{=\sum\limits_{j=0}^{u}(U_b)_{u,j}e_j^T} A = \left(\sum_{j=0}^{u}(U_b)_{u,j}\,e_j^T\right) A = \sum_{j=0}^{u}(U_b)_{u,j}\underbrace{e_j^T A}_{\substack{=e_j^T B \\ \text{(by (16))}}} = \sum_{j=0}^{u}(U_b)_{u,j}\,e_j^T B.$$

Comparing this with

$$\underbrace{e_u^T U_b}_{=\sum\limits_{j=0}^{u}(U_b)_{u,j}e_j^T} B = \left(\sum_{j=0}^{u}(U_b)_{u,j}\,e_j^T\right) B = \sum_{j=0}^{u}(U_b)_{u,j}\,e_j^T B,$$

we obtain $e_u^T U_b A = e_u^T U_b B$.

Forget that we fixed u. We thus have shown that

$$\left(e_u^T U_b A = e_u^T U_b B \qquad \text{for all } u \in \{0,1,\ldots,m-k\}\right).$$

In other words, $U_b A \underset{k}{\equiv} U_b B$ (by the definition of "$U_b A \underset{k}{\equiv} U_b B$"). This proves Lemma 3.10. □

The analogue of Lemma 3.10 for S is even simpler:

Lemma 3.11. Let $\ell \in \mathbb{N} \cup \{\infty\}$. Let $A \in \mathbb{L}^{m\times\ell}$ and $B \in \mathbb{L}^{m\times\ell}$ be two $m\times\ell$-matrices. Let k be a positive integer such that $A \underset{k}{\equiv} B$. Then, $SA \underset{k+1}{\equiv} SB$.

Proof of Lemma 3.11. We have $A \underset{k}{\equiv} B$. In other words, we have

$$\left(e_u^T A = e_u^T B \qquad \text{for all } u \in \{0,1,\ldots,m-k\}\right) \tag{19}$$

(by the definition of "$A \underset{k}{\equiv} B$").

Let $u \in \{0,1,\ldots,m-(k+1)\}$. Then, $u \le m-(k+1) = m-k-1$, so that $u+1 \le m-k$. Combining this with $u+1 \in \mathbb{N}$ (since $u \in \{0,1,\ldots,m-(k+1)\} \subseteq \mathbb{N}$), we obtain $u+1 \in \{0,1,\ldots,m-k\}$. Hence, (19) (applied to $u+1$ instead of u) yields $e_{u+1}^T A = e_{u+1}^T B$.

But $u+1 \le m - \underbrace{k}_{>0} < m$. Hence, Proposition 3.7 **(a)** yields $e_u^T S = e_{u+1}^T$. Hence, $\underbrace{e_u^T S}_{=e_{u+1}^T} A = e_{u+1}^T A = e_{u+1}^T B$. Comparing this with $\underbrace{e_u^T S}_{=e_{u+1}^T} B = e_{u+1}^T B$, we obtain $e_u^T SA = e_u^T SB$.

Forget that we fixed u. We thus have shown that

$$\left(e_u^T SA = e_u^T SB \qquad \text{for all } u \in \{0,1,\ldots,m-(k+1)\}\right).$$

In other words, $SA \underset{k+1}{\equiv} SB$ (by the definition of "$SA \underset{k+1}{\equiv} SB$"). This proves Lemma 3.11. □

Now, we can prove the following lemma, which is as close as we can get to Proposition 3.1 without requiring $m = \infty$:

Lemma 3.12. Let $A \in \mathbb{L}^{m \times 1}$ and $c \in \mathbb{L}$. Let k be a positive integer such that $A \underset{k}{\equiv} H_c$. Then, $SA \underset{k+1}{\equiv} H_{ac}$.

Proof of Lemma 3.12. Lemma 3.11 (applied to $\ell = 1$ and $B = H_c$) yields $SA \underset{k+1}{\equiv} SH_c$. In other words, we have

$$\left(e_u^T SA = e_u^T SH_c \qquad \text{for all } u \in \{0, 1, \ldots, m - (k+1)\}\right) \tag{20}$$

(by the definition of "$SA \underset{k+1}{\equiv} SH_c$").

Now, let $u \in \{0, 1, \ldots, m - (k+1)\}$. Thus, $u \leq m - (k+1) = m - k - 1$, so that $u + 1 \leq m - \underbrace{k}_{>0} < m$. Thus, Proposition 3.7 **(b)** yields $e_u^T SH_c = e_u^T H_{ac}$. But (20) yields

$$e_u^T SA = e_u^T SH_c = e_u^T H_{ac}.$$

Now, forget that we fixed u. We thus have shown that

$$\left(e_u^T SA = e_u^T H_{ac} \qquad \text{for all } u \in \{0, 1, \ldots, m - (k+1)\}\right).$$

In other words, $SA \underset{k+1}{\equiv} H_{ac}$ (by the definition of "$SA \underset{k+1}{\equiv} H_{ac}$"). This proves Lemma 3.12. □

3.7. Proof of Theorem 2.7

Our last stop before Theorem 2.7 is the following lemma, which by now is an easy induction:

Lemma 3.13. Let $b \in \mathbb{L}$. Let $n \in \mathbb{N}$. Then,

$$(U_b S)^n H_1 \underset{n+1}{\equiv} H_{(ba)^n}.$$

Proof of Lemma 3.13. We shall prove Lemma 3.13 by induction on n:

Induction base: It is easy to see that $H_1 \underset{0+1}{\equiv} H_1$ [11].

[11] *Proof.* Clearly,

$$\left(e_u^T H_1 = e_u^T H_1 \qquad \text{for all } u \in \{0, 1, \ldots, m - (0+1)\}\right).$$

In other words, $H_1 \underset{0+1}{\equiv} H_1$ (by the definition of "$H_1 \underset{0+1}{\equiv} H_1$").

But $\underbrace{(U_b S)^0}_{=I_m} H_1 = I_m H_1 = H_1$ and $H_{(ba)^0} = H_1$ (since $(ba)^0 = 1$). In view of these two equalities, we can rewrite $H_1 \underset{0+1}{\equiv} H_1$ as $(U_b S)^0 H_1 \underset{0+1}{\equiv} H_{(ba)^0}$. In other words, Lemma 3.13 holds for $n = 0$. This completes the induction base.

Induction step: Let k be a positive integer. Assume that Lemma 3.13 holds for $n = k - 1$. We must prove that Lemma 3.13 holds for $n = k$.

We have assumed that Lemma 3.13 holds for $n = k - 1$. In other words, we have

$$(U_b S)^{k-1} H_1 \underset{k}{\equiv} H_{(ba)^{k-1}}.$$

Hence, Lemma 3.12 (applied to $A = (U_b S)^{k-1} H_1$ and $c = (ba)^{k-1}$) yields

$$S (U_b S)^{k-1} H_1 \underset{k+1}{\equiv} H_{a(ba)^{k-1}}.$$

Thus, Lemma 3.10 (applied to 1, $k+1$, $S (U_b S)^{k-1} H_1$ and $H_{a(ba)^{k-1}}$ instead of ℓ, k, A and B) yields

$$U_b S (U_b S)^{k-1} H_1 \underset{k+1}{\equiv} U_b H_{a(ba)^{k-1}}.$$

In view of

$$U_b S (U_b S)^{k-1} = (U_b S) (U_b S)^{k-1} = (U_b S)^k$$

and

$$\begin{aligned} U_b H_{a(ba)^{k-1}} &= H_{ba(ba)^{k-1}} && \left(\text{by Proposition 3.8, applied to } c = a (ba)^{k-1}\right) \\ &= H_{(ba)^k} && \left(\text{since } ba (ba)^{k-1} = (ba)(ba)^{k-1} = (ba)^k\right), \end{aligned}$$

this rewrites as

$$(U_b S)^k H_1 \underset{k+1}{\equiv} H_{(ba)^k}.$$

In other words, Lemma 3.13 holds for $n = k$. This completes the induction step. Thus, Lemma 3.13 is proven by induction. □

We can now easily prove Theorem 2.7:

Proof of Theorem 2.7. We have $n < m$, thus $n + 1 \leq m$ (since n and m are integers), hence $m - (n+1) \geq 0$. Thus, $0 \in \{0, 1, \ldots, m - (n+1)\}$. Also, $0 \leq 0 < m$ (since $0 \leq n < m$).

Lemma 3.13 shows that

$$(U_b S)^n H_1 \underset{n+1}{\equiv} H_{(ba)^n}.$$

In other words, we have

$$\left(e_u^T (U_b S)^n H_1 = e_u^T H_{(ba)^n} \qquad \text{for all } u \in \{0, 1, \ldots, m - (n+1)\}\right)$$

(by the definition of "$\left(U_bS\right)^n H_1 \underset{n+1}{\equiv} H_{(ba)^n}$"). We can apply this to $u = 0$ (since $0 \in \{0, 1, \ldots, m - (n+1)\}$), and thus obtain

$$\begin{aligned}
e_0^T \left(U_bS\right)^n H_1 = e_0^T H_{(ba)^n} &= \left(\text{the 0-th row of } H_{(ba)^n}\right) \\
&\qquad \left(\text{by Lemma 3.5, applied to } \ell = 1, i = 0 \text{ and } A = H_{(ba)^n}\right) \\
&= \left(\text{the 0-th entry of } H_{(ba)^n}\right) \qquad \left(\text{since } H_{(ba)^n} \text{ is a column vector}\right) \\
&= \underbrace{a^0}_{=1} (ba)^n \qquad \left(\begin{array}{c}\text{since (4) (applied to } c = (ba)^n\text{)} \\ \text{yields } H_{(ba)^n} = \left(a^i (ba)^n\right)_{0 \le i < m,\ 0 \le j < 1}\end{array}\right) \\
&= (ba)^n .
\end{aligned}$$

This proves Theorem 2.7. □

4. A Weyl-algebraic application

4.1. The claim

We shall now restrict ourselves to a more special situation.

Namely, we let $\mathbb{K}$ be a commutative ring, and we assume that the ring $\mathbb{L}$ is a $\mathbb{K}$-algebra.

Consider the polynomial ring $\mathbb{K}[t]$ in one variable t over $\mathbb{K}$. For each polynomial $g \in \mathbb{K}[t]$ and each $n \in \mathbb{N}$, we let $g^{(n)}$ be the n-th derivative of g; that is,

$$g^{(n)} = \frac{d^n}{dt^n} g. \tag{21}$$

Thus, in particular, $g^{(0)} = g$ and $g^{(1)} = g'$ (where g' denotes the derivative $\frac{d}{dt}g$ of g).

Recall that we fixed $a \in \mathbb{L}$. Furthermore, let $h \in \mathbb{L}$ and $x \in \mathbb{L}$ be such that

$$[a, x] = h \qquad \text{and} \qquad [h, a] = 0 \qquad \text{and} \qquad [h, x] = 0.$$

This situation is actually fairly common:

Example 4.1. Let D be the differentiation operator

$$\mathbb{K}[t] \to \mathbb{K}[t], \qquad g \mapsto \frac{d}{dt} g.$$

Let T be the "multiplication by t" operator

$$\mathbb{K}[t] \to \mathbb{K}[t], \qquad g \mapsto tg.$$

Then, the three operators D, T and $\operatorname{id}_{\mathbb{K}[t]}$ belong to the $\mathbb{K}$-algebra $\operatorname{End}_{\mathbb{K}}(\mathbb{K}[t])$ of all endomorphisms of the $\mathbb{K}$-module $\mathbb{K}[t]$. These three operators satisfy

$$[D,T] = \operatorname{id}_{\mathbb{K}[t]}, \qquad \left[\operatorname{id}_{\mathbb{K}[t]}, D\right] = 0 \qquad \text{and} \qquad \left[\operatorname{id}_{\mathbb{K}[t]}, T\right] = 0.$$

Hence, we can obtain an example of the situation we are considering by setting $\mathbb{L} = \operatorname{End}_{\mathbb{K}}(\mathbb{K}[t])$, $a = D$, $x = T$ and $h = \operatorname{id}_{\mathbb{K}[t]}$.

Further examples can be obtained by varying this one. For example, if $\mathbb{K}$ is a field, then $\mathbb{K}[t]$ can be replaced by the field of rational functions $\mathbb{K}(t)$. Alternatively, if $\mathbb{K} = \mathbb{R}$, then $\mathbb{K}[t]$ can be replaced by the algebra of C^∞-functions $\mathbb{R} \to \mathbb{R}$.

Other examples appear in the theory of Weyl algebras and of 2-step nilpotent Lie algebras.

Now, we return to the generality of $\mathbb{K}$, $\mathbb{L}$, a, h, x and m satisfying $[a,x] = h$ and $[h,a] = 0$ and $[h,x] = 0$.

For any polynomial $g \in \mathbb{K}[t]$, we define an $m \times m$-matrix $V_g \in \mathbb{L}^{m \times m}$ by

$$V_g = \left(\begin{cases} \dbinom{i}{j} g^{(i-j)}(x) \cdot h^{i-j}, & \text{if } i \geq j; \\ 0, & \text{if } i < j \end{cases} \right)_{0 \leq i < m,\ 0 \leq j < m} . \tag{22}$$

This matrix V_g looks as follows:

- If $g \in \mathbb{K}[t]$ and $m \in \mathbb{N}$, then

$$V_g = \begin{pmatrix} g^{(0)}(x) & 0 & 0 & \cdots & 0 \\ g^{(1)}(x) \cdot h & g^{(0)}(x) & 0 & \cdots & 0 \\ g^{(2)}(x) \cdot h^2 & 2g^{(1)}(x) \cdot h & g^{(0)}(x) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g^{(m-1)}(x) \cdot h^{m-1} & (m-1) g^{(m-2)}(x) \cdot h^{m-2} & \binom{m-1}{2} g^{(m-3)}(x) \cdot h^{m-3} & \cdots & g^{(0)}(x) \end{pmatrix}.$$

- If $g \in \mathbb{K}[t]$ and $m = \infty$, then

$$V_g = \begin{pmatrix} g^{(0)}(x) & 0 & 0 & 0 & \cdots \\ g^{(1)}(x) \cdot h & g^{(0)}(x) & 0 & 0 & \cdots \\ g^{(2)}(x) \cdot h^2 & 2g^{(1)}(x) \cdot h & g^{(0)}(x) & 0 & \cdots \\ g^{(3)}(x) \cdot h^3 & 3g^{(2)}(x) \cdot h^2 & 3g^{(1)}(x) \cdot h & g^{(0)}(x) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Now, Tom Copeland has found the following identity [MO337766]:

Theorem 4.2. Let $n \in \mathbb{N}$ be such that $n < m$. Let $g \in \mathbb{K}[t]$. Then,

$$\left(g\left(x\right) \cdot a\right)^n = e_0^T \left(V_g S\right)^n H_1.$$

This identity will easily follow from Theorem 2.7 (applied to $b = g\left(x\right)$), once we can show the following:

Proposition 4.3. Let $g \in \mathbb{K}[t]$. Then, $U_{g(x)} = V_g$.

We shall thus mostly focus on proving Proposition 4.3.

4.2. How derivatives appear in commutators

The main idea of our proof will be the following proposition, which relates derivatives in $\mathbb{K}[t]$ to commutators in $\mathbb{L}$:

Proposition 4.4. (a) We have $ax^i = x^i a + ix^{i-1}h$ for each positive integer i.
(b) We have $a \cdot g\left(x\right) = g\left(x\right) \cdot a + g'\left(x\right) \cdot h$ for each $g \in \mathbb{K}[t]$.

Here, of course, g' means the derivative $\dfrac{d}{dt}g$ of the polynomial g.

Proof of Proposition 4.4. The definition of $[a, x]$ yields $[a, x] = ax - xa$. Hence, $ax - xa = [a, x] = h$. Thus, $ax = xa + h$.

From $[h, x] = 0$, we obtain $0 = [h, x] = hx - xh$ (by the definition of $[h, x]$). In other words, $hx = xh$.

(a) We shall prove Proposition 4.4 **(a)** by induction on i:

Induction base: Comparing $a \underbrace{x^1}_{=x} = ax = xa + h$ with $\underbrace{x^1}_{=x} a + 1 \underbrace{x^{1-1}}_{=x^0=1} h = xa + h$, we find $ax^1 = x^1 a + 1x^{1-1}h$. In other words, Proposition 4.4 **(a)** holds for $i = 1$. This completes the induction base.

Induction step: Let n be a positive integer. Assume that Proposition 4.4 **(a)** holds for $i = n$. We must prove that Proposition 4.4 **(a)** holds for $i = n + 1$.

We have assumed that Proposition 4.4 **(a)** holds for $i = n$. In other words, we have $ax^n = x^n a + nx^{n-1}h$.

Now,

$$\begin{aligned} a \underbrace{x^{n+1}}_{=x^n x} &= \underbrace{ax^n}_{=x^n a + nx^{n-1}h} x = \left(x^n a + nx^{n-1}h\right) x = x^n \underbrace{ax}_{=xa+h} + nx^{n-1} \underbrace{hx}_{=xh} \\ &= \underbrace{x^n \left(xa + h\right)}_{=x^n xa + x^n h} + n \underbrace{x^{n-1} x}_{=x^n} h = \underbrace{x^n x}_{=x^{n+1}} a + \underbrace{x^n h + nx^n h}_{=(n+1)x^n h} \\ &= x^{n+1} a + \left(n+1\right) \underbrace{x^n}_{=x^{(n+1)-1}} h = x^{n+1} a + \left(n+1\right) x^{(n+1)-1} h. \end{aligned}$$

In other words, Proposition 4.4 **(a)** holds for $i = n+1$. This completes the induction step. Thus, Proposition 4.4 **(a)** is proven by induction.

(b) Let $g \in \mathbb{K}[t]$. Write the polynomial g in the form $g = \sum\limits_{i=0}^{k} g_i t^i$ for some $k \in \mathbb{N}$ and some $g_0, g_1, \ldots, g_k \in \mathbb{K}$. Thus, the definition of the derivative g' yields $g' = \sum\limits_{i=1}^{k} i g_i t^{i-1}$. Substituting x for t in this equality, we find

$$g'(x) = \sum_{i=1}^{k} \underbrace{i g_i}_{= g_i i} x^{i-1} = \sum_{i=1}^{k} g_i i x^{i-1}. \tag{23}$$

Substituting x for t in the equality $g = \sum\limits_{i=0}^{k} g_i t^i$, we obtain $g(x) = \sum\limits_{i=0}^{k} g_i x^i$. Hence,

$$\begin{aligned} a \cdot \underbrace{g(x)}_{= \sum\limits_{i=0}^{k} g_i x^i} &= a \cdot \sum_{i=0}^{k} g_i x^i = \sum_{i=0}^{k} g_i a x^i = g_0 a \underbrace{x^0}_{=1} + \sum_{i=1}^{k} g_i \underbrace{a x^i}_{\substack{= x^i a + i x^{i-1} h \\ \text{(by Proposition 4.4 (a))}}} \\ &\qquad \text{(here, we have split off the addend for } i = 0 \text{ from the sum)} \\ &= g_0 a + \underbrace{\sum_{i=1}^{k} g_i \left(x^i a + i x^{i-1} h \right)}_{= \sum\limits_{i=1}^{k} g_i x^i a + \sum\limits_{i=1}^{k} g_i i x^{i-1} h} = g_0 a + \sum_{i=1}^{k} g_i x^i a + \sum_{i=1}^{k} g_i i x^{i-1} h. \end{aligned}$$

Comparing this with

$$\begin{aligned} &\underbrace{g(x)}_{= \sum\limits_{i=0}^{k} g_i x^i} \cdot a + \underbrace{g'(x)}_{\substack{= \sum\limits_{i=1}^{k} g_i i x^{i-1} \\ \text{(by (23))}}} \cdot h \\ &= \left(\sum_{i=0}^{k} g_i x^i \right) \cdot a + \left(\sum_{i=1}^{k} g_i i x^{i-1} \right) \cdot h = \underbrace{\sum_{i=0}^{k} g_i x^i a}_{\substack{= g_0 x^0 a + \sum\limits_{i=1}^{k} g_i x^i a \\ \text{(here, we have split off the} \\ \text{addend for } i=0 \text{ from the sum)}}} + \sum_{i=1}^{k} g_i i x^{i-1} h \\ &= g_0 \underbrace{x^0}_{=1} a + \sum_{i=1}^{k} g_i x^i a + \sum_{i=1}^{k} g_i i x^{i-1} h = g_0 a + \sum_{i=1}^{k} g_i x^i a + \sum_{i=1}^{k} g_i i x^{i-1} h, \end{aligned}$$

we obtain $a \cdot g(x) = g(x) \cdot a + g'(x) \cdot h$. This proves Proposition 4.4 **(b)**. □

Note that we have not used the condition $[h, a] = 0$ in Proposition 4.4; but we will use it now:

Proposition 4.5. Let $b \in \mathbb{L}$. Then, $\operatorname{ad}_a\left(bh^i\right) = \operatorname{ad}_a(b) \cdot h^i$ for each $i \in \mathbb{N}$.

Proof of Proposition 4.5. From $[h,a] = 0$, we obtain $0 = [h,a] = ha - ah$ (by the definition of $[h,a]$). In other words, $ha = ah$. Hence,

$$h^i a = ah^i \qquad \text{for each } i \in \mathbb{N}. \tag{24}$$

[*Proof of (24):* This follows by induction on i, using $ha = ah$ in the induction step.]

Now, let $i \in \mathbb{N}$. Then, the definition of ad_a yields

$$\operatorname{ad}_a(b) = [a,b] = ab - ba \qquad (\text{by the definition of } [a,b]).$$

But the definition of ad_a also yields

$$\begin{aligned}\operatorname{ad}_a\left(bh^i\right) &= \left[a, bh^i\right] = a\left(bh^i\right) - \left(bh^i\right)a \qquad \left(\text{by the definition of } \left[a, bh^i\right]\right)\\ &= abh^i - b\underbrace{h^i a}_{\substack{=ah^i\\ \text{(by (24))}}} = abh^i - bah^i = \underbrace{(ab - ba)}_{=\operatorname{ad}_a(b)} h^i\\ &= \operatorname{ad}_a(b) \cdot h^i.\end{aligned}$$

This proves Proposition 4.5. □

Corollary 4.6. Let $g \in \mathbb{K}[t]$. Let $p \in \mathbb{N}$. Then,

$$\operatorname{ad}_a^p(g(x)) = g^{(p)}(x) \cdot h^p.$$

Proof of Corollary 4.6. We shall prove Corollary 4.6 by induction on p:

Induction base: Comparing $\underbrace{\operatorname{ad}_a^0}_{=\operatorname{id}}(g(x)) = \operatorname{id}(g(x)) = g(x)$ with $\underbrace{g^{(0)}}_{=g}(x) \cdot \underbrace{h^0}_{=1} = g(x)$, we obtain $\operatorname{ad}_a^0(g(x)) = g^{(0)}(x) \cdot h^0$. In other words, Corollary 4.6 holds for $p = 0$. This completes the induction base.

Induction step: Let $n \in \mathbb{N}$. Assume that Corollary 4.6 holds for $p = n$. We must prove that Corollary 4.6 holds for $p = n + 1$.

We have assumed that Corollary 4.6 holds for $p = n$. In other words, we have

$$\operatorname{ad}_a^n(g(x)) = g^{(n)}(x) \cdot h^n.$$

Now,

$$\begin{aligned}\underbrace{\operatorname{ad}_a^{n+1}}_{=\operatorname{ad}_a \circ \operatorname{ad}_a^n}(g(x)) &= \left(\operatorname{ad}_a \circ \operatorname{ad}_a^n\right)(g(x)) = \operatorname{ad}_a\left(\underbrace{\operatorname{ad}_a^n(g(x))}_{=g^{(n)}(x)\cdot h^n}\right)\\ &= \operatorname{ad}_a\left(g^{(n)}(x) \cdot h^n\right) = \operatorname{ad}_a\left(g^{(n)}(x)\right) \cdot h^n \end{aligned} \tag{25}$$

(by Proposition 4.5, applied to $b = g^{(n)}(x)$ and $i = n$).

But Proposition 4.4 **(b)** (applied to $g^{(n)}$ instead of g) yields

$$a \cdot g^{(n)}(x) = g^{(n)}(x) \cdot a + \left(g^{(n)}\right)'(x) \cdot h.$$

In view of $\left(g^{(n)}\right)' = g^{(n+1)}$ (this follows from the definitions of $g^{(n)}$ and $g^{(n+1)}$), we can rewrite this as

$$a \cdot g^{(n)}(x) = g^{(n)}(x) \cdot a + g^{(n+1)}(x) \cdot h. \tag{26}$$

Now, the definition of ad_a yields

$$\begin{aligned}\operatorname{ad}_a\left(g^{(n)}(x)\right) &= \left[a, g^{(n)}(x)\right] = a \cdot g^{(n)}(x) - g^{(n)}(x) \cdot a \\ &\qquad \left(\text{by the definition of } \left[a, g^{(n)}(x)\right]\right) \\ &= g^{(n+1)}(x) \cdot h \qquad \text{(by (26))}.\end{aligned}$$

Hence, (25) becomes

$$\operatorname{ad}_a^{n+1}(g(x)) = \underbrace{\operatorname{ad}_a\left(g^{(n)}(x)\right)}_{=g^{(n+1)}(x)\cdot h} \cdot h^n = g^{(n+1)}(x) \cdot \underbrace{h \cdot h^n}_{=h^{n+1}} = g^{(n+1)}(x) \cdot h^{n+1}.$$

In other words, Corollary 4.6 holds for $p = n+1$. This completes the induction step. Thus, Corollary 4.6 is proven by induction. □

4.3. Proofs of Proposition 4.3 and Theorem 4.2

Proof of Proposition 4.3. If $b \in \mathbb{L}$ is arbitrary, then

$$(U_b)_{i,j} = \begin{cases} \dbinom{i}{j} \operatorname{ad}_a^{i-j}(b), & \text{if } i \geq j; \\ 0, & \text{if } i < j \end{cases} \tag{27}$$

for each $i \in \{0, 1, \ldots, m-1\}$ and $j \in \{0, 1, \ldots, m-1\}$ (by (3)).

On the other hand, (22) shows that

$$\left(V_g\right)_{i,j} = \begin{cases} \dbinom{i}{j} g^{(i-j)}(x) \cdot h^{i-j}, & \text{if } i \geq j; \\ 0, & \text{if } i < j \end{cases} \tag{28}$$

for each $i \in \{0, 1, \ldots, m-1\}$ and $j \in \{0, 1, \ldots, m-1\}$.

Now, let us fix $i \in \{0, 1, \ldots, m-1\}$ and $j \in \{0, 1, \ldots, m-1\}$. We shall prove that $\left(U_{g(x)}\right)_{i,j} = \left(V_g\right)_{i,j}$.

Indeed, we are in one of the following two cases:

Case 1: We have $i \geq j$.

Case 2: We have $i < j$.

Let us first consider Case 1. In this case, we have $i \geq j$. Hence, $i - j \in \mathbb{N}$. Thus, Corollary 4.6 (applied to $p = i - j$) yields

$$\operatorname{ad}_a^{i-j}(g(x)) = g^{(i-j)}(x) \cdot h^{i-j}.$$

Now, (27) (applied to $b = g(x)$) yields

$$\begin{aligned}\left(U_{g(x)}\right)_{i,j} &= \begin{cases} \dbinom{i}{j} \operatorname{ad}_a^{i-j}(g(x)), & \text{if } i \geq j; \\ 0, & \text{if } i < j \end{cases} = \dbinom{i}{j} \underbrace{\operatorname{ad}_a^{i-j}(g(x))}_{= g^{(i-j)}(x) \cdot h^{i-j}} \qquad (\text{since } i \geq j) \\ &= \dbinom{i}{j} g^{(i-j)}(x) \cdot h^{i-j}.\end{aligned}$$

Comparing this with

$$\begin{aligned}(V_g)_{i,j} &= \begin{cases} \dbinom{i}{j} g^{(i-j)}(x) \cdot h^{i-j}, & \text{if } i \geq j; \\ 0, & \text{if } i < j \end{cases} \qquad (\text{by (28)}) \\ &= \dbinom{i}{j} g^{(i-j)}(x) \cdot h^{i-j} \qquad (\text{since } i \geq j),\end{aligned}$$

we obtain $\left(U_{g(x)}\right)_{i,j} = (V_g)_{i,j}$. Thus, $\left(U_{g(x)}\right)_{i,j} = (V_g)_{i,j}$ is proven in Case 1.

Let us next consider Case 2. In this case, we have $i < j$. Hence, (28) becomes

$$(V_g)_{i,j} = \begin{cases} \dbinom{i}{j} g^{(i-j)}(x) \cdot h^{i-j}, & \text{if } i \geq j; \\ 0, & \text{if } i < j \end{cases} = 0 \qquad (\text{since } i < j).$$

But (27) (applied to $b = g(x)$) yields

$$\left(U_{g(x)}\right)_{i,j} = \begin{cases} \dbinom{i}{j} \operatorname{ad}_a^{i-j}(g(x)), & \text{if } i \geq j; \\ 0, & \text{if } i < j \end{cases} = 0 \qquad (\text{since } i < j).$$

Comparing these two equalities, we find $\left(U_{g(x)}\right)_{i,j} = (V_g)_{i,j}$. Thus, $\left(U_{g(x)}\right)_{i,j} = (V_g)_{i,j}$ is proven in Case 2.

We have now proven the equality $\left(U_{g(x)}\right)_{i,j} = (V_g)_{i,j}$ in both Cases 1 and 2. Hence, this equality always holds.

Now, forget that we fixed i and j. We thus have shown that $\left(U_{g(x)}\right)_{i,j} = \left(V_g\right)_{i,j}$ for all $i \in \{0, 1, \ldots, m-1\}$ and $j \in \{0, 1, \ldots, m-1\}$. In other words, each entry of the $m \times m$-matrix $U_{g(x)}$ equals the corresponding entry of the $m \times m$-matrix V_g. Hence, $U_{g(x)} = V_g$. This proves Proposition 4.3. □

Proof of Theorem 4.2. Theorem 2.7 (applied to $b = g(x)$) yields

$$(g(x) \cdot a)^n = e_0^T \left(\underbrace{U_{g(x)}}_{\substack{=V_g \\ \text{(by Proposition 4.3)}}} S \right)^n H_1 = e_0^T \left(V_g S\right)^n H_1.$$

This proves Theorem 4.2. □

References

[MO337766] Tom Copeland, *MathOverflow question #337766 ("Expansions of iterated, or nested, derivatives, or vectors – conjectured matrix computation").* `https://mathoverflow.net/questions/337766/`