



**HAL**  
open science

## Growth of structure constants of free Lie algebras relative to Hall bases

Karine Beauchard, Jérémy Le Borgne, Frédéric Marbach

► **To cite this version:**

Karine Beauchard, Jérémy Le Borgne, Frédéric Marbach. Growth of structure constants of free Lie algebras relative to Hall bases. *Journal of Algebra*, 2022, 612, pp.281-378. 10.1016/j.jalgebra.2022.08.030 . hal-03451998v3

**HAL Id: hal-03451998**

**<https://hal.science/hal-03451998v3>**

Submitted on 30 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Growth of structure constants of free Lie algebras relative to Hall bases

Karine Beauchard\*, Jérémy Le Borgne\*, Frédéric Marbach\*

August 30, 2022

## Abstract

We derive *a priori* bounds on the size of the structure constants of the free Lie algebra over a set of indeterminates, relative to its Hall bases. We investigate their asymptotic growth, especially as a function of the length of the involved Lie brackets.

First, using the classical recursive decomposition algorithm, we obtain a rough upper bound valid for all Hall bases. We then introduce new notions (which we call *alphabetic subsets* and *relative foldings*) related to structural properties of the Lie brackets created by the algorithm, which allow us to prove a sharp upper bound for the general case. We also prove that the length of the relative folding provides a strictly decreasing indexation of the recursive rewriting algorithm. Moreover, we derive lower bounds on the structure constants proving that they grow at least geometrically in all Hall bases.

Second, for the celebrated historical length-compatible Hall bases and the Lyndon basis, we prove tighter sharp upper bounds, which turn out to be geometric in the length of the brackets.

Third, we construct two new Hall bases, illustrating two opposite behaviors in the two-indeterminates case. One is designed so that its structure constants have the minimal growth possible, matching exactly the general lower bound, linked with the Fibonacci sequence. The other one is designed so that its structure constants grow super-geometrically.

Eventually, we investigate asymmetric growth bounds which isolate the role of one particular indeterminate. Despite the existence of super-geometric Hall bases, we prove that the asymmetric growth with respect to each fixed indeterminate is uniformly at most geometric in all Hall bases.

---

\*Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France

# Contents

1	Introduction	3
2	A rough bound for the decomposition algorithm	10
3	Iterated brackets over alphabetic subsets	13
4	A refined bound stemming from brackets structure	17
5	Systematic geometric lower bounds	26
6	Length-compatible Hall sets	30
7	Lyndon basis	35
8	A minimal Hall set	39
9	A super-geometric Hall set	55
10	Asymmetric estimates	59
11	Perspectives and open problems	65
A	Detailed example of the decomposition algorithm	66
B	A Leibniz-type inversion rule	66
C	Computations and estimates on some binomial sums	67
D	Some estimates with Fibonacci numbers	70
E	Other bases of free Lie algebras	71

# 1 Introduction

## 1.1 Context and motivation

We consider Hall bases of the free Lie algebra  $\mathcal{L}(X)$  over a set  $X$  with  $|X| \geq 2$ . This class of bases, understood in the generalized sense of Viennot [48, Theorem 1.2] or Shirshov [39, Definition 1], includes many well-known and widely used bases of  $\mathcal{L}(X)$  (see Section 1.4 for more details).

Given a Hall set  $\mathcal{B} \subset \text{Br}(X)$ , the free magma over  $X$ , we investigate the growth of the structure constants of  $\mathcal{L}(X)$  relative to the Hall basis associated with  $\mathcal{B}$ . More precisely, we estimate the  $\ell^1$  norm of the Lie bracket  $[a, b]$  of two elements  $a$  and  $b$  of the basis, denoted by  $\|[a, b]\|_{\mathcal{B}}$ , and we track its growth with respect to the size of the involved basis elements. We investigate both *symmetric estimates*, where the growth is measured with respect to the usual lengths  $|a|$  and  $|b|$  of the involved elements of the basis, and *asymmetric estimates*, where the growth is tracked with a special focus on the number of occurrences of a singled-out indeterminate  $X_0 \in X$ .

To the best of our knowledge, this work is the first providing estimates on the structure constants of free Lie algebras. Classical mathematical questions around structure constants involve their computation for particular algebras (see e.g. [36] for spherical Hecke algebras, [13] for symmetric Hopf algebras, [1] for a superoperator algebra, [6] for Kac-Moody Lie algebras, [7, 9, 11] for semisimple Lie algebras), the investigation of their sign [14, 47], or the identification and reconstruction of the algebra from them [31, 45].

Our initial motivation, both for symmetric and asymmetric estimates, stems from convergence issues for series of Lie brackets of analytic vector fields (see e.g. the open problem raised in [2, Section 2.4.3] and the conditional result in [2, Section 4.4.3]). From an algebraic point of view, such estimates are linked with the intricacies of the Lie product in a Hall basis. In particular, the methods we develop could probably be extended to estimate the computational complexity (with respect to time or space) of algorithms used by Lie algebraic packages to decompose Lie brackets of the free Lie algebra on a Hall basis (see Section 2.1 for more details).

## 1.2 Main results and plan of the paper

In this paragraph, we state rough pedagogical versions of our main results and sketch the plan of the paper. We refer to Section 1.3 for definitions of the algebraic notions involved, to Section 1.4 for definitions and properties relative to Hall sets and our choice of norm relative to a basis (see in particular (1.18)), and to the sequel of the paper for enhanced statements of the results.

First, in Section 2, using the classical recursive algorithm which allows to decompose a Lie bracket  $[a, b]$  of two basis elements on the Hall basis (see e.g. [33, Section 9]), we obtain the following very rough bound.

**Theorem 1.1.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. For every  $a < b \in \mathcal{B}$ ,*

$$\|[a, b]\|_{\mathcal{B}} \leq 2^{\binom{n(n+1)}{2}} \quad \text{where } n := |a| + |b|. \quad (1.1)$$

Then, in Section 3, we introduce the notion of *alphabetic subsets* of  $\mathcal{B}$ . As a prime example, the set  $X \subset \mathcal{B}$  is an alphabetic subset. This notion, along with an associated stability property, allows us to obtain the following upper bound on iterated Lie brackets of elements of such subsets, which is sharp in the sense of Corollary 3.13.

**Theorem 1.2.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. Let  $A$  be an alphabetic subset of  $\mathcal{B}$ . For every  $n \geq 2$ , and  $t$  an iterated Lie bracket of  $n$  elements of  $A$ ,*

$$\|t\|_{\mathcal{B}} \leq (n-1)!. \quad (1.2)$$

In Section 4, we introduce the notion of *relative folding* of an element  $b \in \mathcal{B}$  with respect to another element  $a \in \mathcal{B}$  such that  $a < b$ . This construction, along with the key remark that it yields an iterated Lie bracket over an alphabetic subset, allows us to describe the elements of  $\mathcal{B}$  involved in the decomposition of  $[a, b]$  and prove the following bound, which is sharp in the sense of Corollary 4.15. We also prove that the length of the relative folding provides a nice strictly decreasing indexation for the classical recursive rewriting algorithm (see Section 4.5).

**Theorem 1.3.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. For every  $a < b \in \mathcal{B}$ ,*

$$\|[a, b]\|_{\mathcal{B}} \leq \lfloor e(|b| - 1)! \rfloor. \quad (1.3)$$

Concerning lower bounds in the general case, we prove in Section 5 that the growth is always at least geometric in all Hall bases, with a common ratio depending on  $|X|$ .

**Theorem 1.4.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. For every  $n \geq 3$ , there exist  $a < b \in \mathcal{B}$  such that  $|a| = 1$ ,  $|b| = n - 1$  (so  $|a| + |b| = n$ ) and*

$$\|[a, b]\|_{\mathcal{B}} \geq \begin{cases} 2^{n-2} & \text{when } |X| \geq 3, \\ F_{n-2} & \text{when } |X| = 2, \end{cases} \quad (1.4)$$

where  $(F_\nu)_{\nu \in \mathbb{N}}$  denote the Fibonacci numbers.

We then turn to estimates specific to particular well-known Hall bases. In Section 6, for the case of the historical Hall bases for which the order is compatible with the length of the brackets, we obtain the following geometric bound, which is sharp in the sense of Proposition 6.11.

**Theorem 1.5.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a length-compatible Hall set. For every  $a < b \in \mathcal{B}$ ,*

$$\|[a, b]\|_{\mathcal{B}} \leq 2^{|b|-1}. \quad (1.5)$$

In Section 7, we obtain a geometric bound for the celebrated Lyndon basis, which is sharp in the sense of Proposition 7.6.

**Theorem 1.6.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be the Hall set of the Lyndon basis. For every  $a < b \in \mathcal{B}$ ,*

$$\|[a, b]\|_{\mathcal{B}} \leq 2^{|b|-1}. \quad (1.6)$$

When  $|X| \geq 3$ , the lower bound of (1.4) is optimal since it matches the upper bounds for the length-compatible and Lyndon bases. When  $|X| = 2$ , we construct in Section 8 an example of a Hall set of “minimal worst product size” in the sense that the following result holds.

**Theorem 1.7.** *When  $|X| = 2$ , there exists a Hall set  $\mathcal{B} \subset \text{Br}(X)$  such that for every  $a < b \in \mathcal{B}$  with  $n := |a| + |b| \geq 3$ ,*

$$\|[a, b]\|_{\mathcal{B}} \leq F_{n-2}. \quad (1.7)$$

In the opposite direction, we construct in Section 9 a particular Hall set which shows that, contrary to the three previous cases, the growth can be super-geometric. Our construction holds even when  $|X| = 2$ .

**Theorem 1.8.** *There exists a Hall set  $\mathcal{B} \subset \text{Br}(X)$  such that, for every  $M \geq 1$ , for every  $n \in \mathbb{N}^*$  large enough, there exist  $a < b \in \mathcal{B}$  with  $|a| = 2$  and  $|b| = n$  such that*

$$\|[a, b]\|_{\mathcal{B}} \geq M^n. \quad (1.8)$$

Eventually, we investigate in Section 10 what we call *asymmetric estimates*, i.e. estimates where we single out the role of a particular indeterminate  $X_0 \in X$ . For  $b \in \text{Br}(X)$ , we denote by  $n_0(b)$  the number of occurrences of  $X_0$  in  $b$ , and  $n(b) := |b| - n_0(b)$  the number of leaves of  $b$  that are different from  $X_0$ . We seek estimates on the structure constants where we attempt to have the lowest asymptotic growth with respect to  $n_0$ . Our main result is the following.

**Theorem 1.9.** *There exists a (universal) sequence  $(C(n))_{n \in \mathbb{N}} \in \mathbb{N}^{\mathbb{N}}$ , such that, if  $\mathcal{B} \subset \text{Br}(X)$  is a Hall set and  $X_0 \in X$ , then, for every  $a < b \in \mathcal{B}$ ,*

$$\|[a, b]\|_{\mathcal{B}} \leq C(n(b))^{n_0(b)} (n(b))!. \quad (1.9)$$

### 1.3 Usual algebraic structures and notations

Let  $\mathbb{K} = \mathbb{R}$  or  $\mathbb{C}$ . Implicitly, all vector spaces and algebras are constructed over the base field  $\mathbb{K}$ . Here and in the sequel,  $X$  is a set, containing at least two elements, not necessarily finite.

#### 1.3.1 Free magma and free monoid

For more detailed constructions of free magmas and monoids, we refer to [4, Chapter 1, §7].

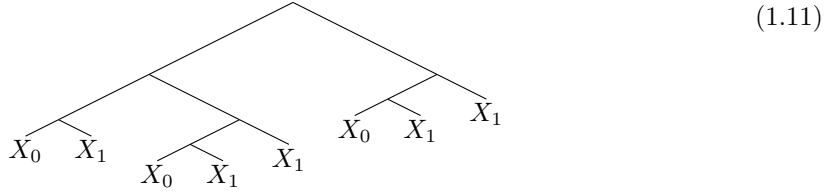
**Definition 1.10** (Free magma). *We consider  $\text{Br}(X)$  the free magma over  $X$ . This set can be defined by induction: for  $X_i \in X$ ,  $X_i \in \text{Br}(X)$  and if  $t_1, t_2 \in \text{Br}(X)$ , then the ordered pair  $(t_1, t_2)$  belongs to  $\text{Br}(X)$ . Intuitively,  $\text{Br}(X)$  can be seen as the set of formal brackets of elements of  $X$ . It is also the set of rooted binary trees, with leaves labeled by elements of  $X$ .*

**Definition 1.11** (Length, left and right factors). *For  $t \in \text{Br}(X)$ ,  $|t|$  denotes the length of  $t$  i.e. the number of leaves of the tree. If  $|t| > 1$ ,  $t$  can be written in a unique way as  $t = (t_1, t_2)$ , with  $t_1, t_2 \in \text{Br}(X)$ . We use the notations  $\lambda(t) = t_1$  and  $\mu(t) = t_2$ , which define maps  $\lambda, \mu : \text{Br}(X) \setminus X \rightarrow \text{Br}(X)$ .*

**Definition 1.12** (Set of iterated left factors). *For  $t \in \text{Br}(X)$ , define*

$$\Lambda(t) := \{\lambda^k(t); k \in \mathbb{N} \text{ such that } |\lambda^k(t)| \geq 1\} \subset \text{Br}(X). \quad (1.10)$$

**Example 1.13.** *The element  $t := (((X_0, X_1), ((X_0, X_1), X_1)), ((X_0, X_1), X_1))$  of  $\text{Br}(\{X_0, X_1\})$  can be visualized as the following tree:*



Here,  $|t| = 8$ ,  $\lambda(t) = ((X_0, X_1), ((X_0, X_1), X_1))$  and  $\mu(t) = ((X_0, X_1), X_1)$ . Moreover,  $\Lambda(t) = \{t, \lambda(t), (X_0, X_1), X_0\}$ .

**Definition 1.14** (Iterated bracketing). *For  $t_1, t_2 \in \text{Br}(X)$ , we introduce  $\text{ad}_{t_1}(t_2) := (t_1, t_2)$ , the left bracketing by  $t_1$  and  $\underline{\text{ad}}_{t_1}(t_2) := (t_2, t_1)$ , the right bracketing by  $t_1$ , which allow us to write compactly iterated brackets (e.g.  $((t_2, t_1), t_1), t_1) = \underline{\text{ad}}_{t_1}^3(t_2)$ ).*

**Definition 1.15** (Free monoid). *We denote by  $X^*$  the free monoid over  $X$ . It is the set of finite sequences of elements of  $X$ , endowed with the concatenation operation. It can be thought of as the set of words over the alphabet whose letters are the elements of  $X$ .*

#### 1.3.2 Higher order brackets

**Definition 1.16** (Higher order brackets). *For  $A \subset \text{Br}(X)$ ,  $\text{Br}(A)$  denotes the free magma over  $A$ . Thus  $\text{Br}(A)$  is a set of formal brackets of formal brackets, i.e. a subset of  $\text{Br}(\text{Br}(X))$ . By convention, the brackets in  $\text{Br}(A)$  are denoted by  $\langle \cdot, \cdot \rangle$  to distinguish them from the brackets  $(\cdot, \cdot)$  in  $\text{Br}(X)$ . We denote by  $\mathbf{I}$  the canonical morphism of magmas from  $\text{Br}(\text{Br}(X))$  to  $\text{Br}(X)$ . We also denote by  $\mathbf{I}$  its restriction to  $\text{Br}(A)$ . For  $t \in \text{Br}(A)$ ,  $|t|_A$  denotes its length in  $\text{Br}(A)$  and  $|\mathbf{I}(t)|$  its length in  $\text{Br}(X)$ .*

**Example 1.17.** *If  $X = \{X_1, X_2, X_3\}$  and  $A = \{(X_1, X_2), X_3\}$ , then  $t = \langle X_3, \langle X_3, (X_1, X_2) \rangle \rangle \in \text{Br}(A)$ ,  $|t|_A = 3$ ,  $\mathbf{I}(t) = (X_3, (X_3, (X_1, X_2)))$  and  $|\mathbf{I}(t)| = 4$ .*

**Definition 1.18** (Submagma generated by a subset). *Let  $A \subset \text{Br}(X)$ . We denote by  $\text{Br}_A \subset \text{Br}(X)$  the submagma of  $\text{Br}(X)$  generated by  $A$ , i.e. the image of the canonical morphism of magmas  $\mathfrak{I} : \text{Br}(A) \rightarrow \text{Br}(X)$ .*

**Lemma 1.19** (Free subset of  $\text{Br}(X)$ ). *Let  $A \subset \text{Br}(X)$ . We say that the subset  $A$  is free when  $A \cap (\text{Br}_A, \text{Br}_A) = \emptyset$ . Then the canonical surjection  $\mathfrak{I} : \text{Br}(A) \rightarrow \text{Br}_A$  is an isomorphism.*

*Proof.* By contradiction, assume that there exists a couple  $t_1 \neq t_2 \in \text{Br}(A)$  such that  $\mathfrak{I}(t_1) = \mathfrak{I}(t_2)$ . If  $|t_1|_A > 1$  and  $|t_2|_A > 1$ , since  $t_1 \neq t_2$ , we can assume by symmetry that  $\lambda(t_1) \neq \lambda(t_2)$ . Since  $\mathfrak{I}(t_1) = \mathfrak{I}(t_2)$ ,  $\mathfrak{I}(\lambda(t_1)) = \mathfrak{I}(\lambda(t_2))$ , so we have found a couple with the same property but shorter lengths in  $\text{Br}(A)$ . Hence, we can assume that  $|t_1|_A = 1$  or  $|t_2|_A = 1$ . One cannot have  $|t_1|_A = |t_2|_A = 1$ , because the canonical surjection is the identity for trees which are a single leaf. By symmetry, assume that  $|t_1|_A = 1$  but  $|t_2|_A > 1$ , then  $\mathfrak{I}(t_1) \in A$  and  $\mathfrak{I}(t_2) \in (\text{Br}_A, \text{Br}_A)$ , which contradicts the assumption that  $A$  is free.  $\square$

### 1.3.3 Free Lie algebra

**Definition 1.20** (Free algebra). *We consider  $\mathcal{A}(X)$  the free associative algebra generated by  $X$  over the field  $\mathbb{K}$ , i.e. the unital associative algebra of polynomials of the noncommutative indeterminates  $X$  (see also [4, Chapter 3, Section 2.7, Definition 2]). The algebra  $\mathcal{A}(X)$  is the free vector space over  $X^*$ , and therefore is a graded algebra:*

$$\mathcal{A}(X) = \bigoplus_{n \in \mathbb{N}} \mathcal{A}_n(X), \quad (1.12)$$

where  $\mathcal{A}_n(X)$  is the finite-dimensional  $\mathbb{K}$ -vector space spanned by monomials of degree  $n$  over  $X$  (i.e. elements of  $X^*$  of length  $n$ ). In particular  $\mathcal{A}_0(X) = \mathbb{K}$  and  $\mathcal{A}_1(X) = \text{span}_{\mathbb{K}}(X)$ .

Let us now recall the definition of the main objects that we will be interested in in this paper: free Lie algebras. We refer to the books [22, 32, 37] and the essay [8] for thorough introductions to Lie algebras and free Lie algebras.

**Definition 1.21** (Free Lie algebra). *The algebra  $\mathcal{A}(X)$  is endowed with a natural structure of Lie algebra, the Lie bracket operation being defined by  $[a, b] := ab - ba$ . This operation satisfies  $[a, a] = 0$  and the Jacobi identity  $[a, [b, c]] + [c, [a, b]] + [b, [c, a]] = 0$ . We consider  $\mathcal{L}(X)$ , the free Lie algebra generated by  $X$  over the field  $\mathbb{K}$ , which is defined as the Lie subalgebra generated by  $X$  in  $\mathcal{A}(X)$ . It can be seen as the smallest linear subspace of  $\mathcal{A}(X)$  containing all elements of  $X$  and stable by the Lie bracket (see also [32, Theorem 0.4]). The Lie algebra  $\mathcal{L}(X)$  is a graded Lie algebra:*

$$\mathcal{L}(X) = \bigoplus_{n \in \mathbb{N}} \mathcal{L}_n(X), \quad [\mathcal{L}_m(X), \mathcal{L}_n(X)] \subset \mathcal{L}_{m+n}(X) \quad (1.13)$$

where, for each  $n \in \mathbb{N}$ , we define  $\mathcal{L}_n(X) := \mathcal{L}(X) \cap \mathcal{A}_n(X)$

**Definition 1.22** (Natural evaluation). *By universal property of the free magma  $\text{Br}(X)$ , there is an ‘‘evaluation’’ mapping  $\mathfrak{E} : \text{Br}(X) \rightarrow \mathcal{L}(X)$ . It is such that  $\mathfrak{E}(X_i) := X_i$  for  $X_i \in X$ , and  $\mathfrak{E}(t) := [\mathfrak{E}(\lambda(t)), \mathfrak{E}(\mu(t))]$  when  $t \in \text{Br}(X) \setminus X$ .*

**Remark 1.23.** *The evaluation map  $\mathfrak{E}$  is not injective: for example,  $(X_0, X_0)$  and  $(X_0, (X_1, X_1))$  are two different elements of  $\text{Br}(X)$ , both evaluated to zero in  $\mathcal{L}(X)$ .*

**Remark 1.24** (Implicit evaluations). *For the sake of readability, we will sometimes, when no confusion is possible, omit the evaluation  $\mathfrak{E}$  when writing equalities in  $\mathcal{L}(X)$ . In particular, since, by convention, we use parentheses  $(\cdot, \cdot)$  as brackets in  $\text{Br}(X)$  and  $[\cdot, \cdot]$  as brackets in  $\mathcal{L}(X)$ , we will simply write  $[a, b]$  as a shorthand for  $[\mathfrak{E}(a), \mathfrak{E}(b)] = \mathfrak{E}((a, b))$  when  $a, b \in \text{Br}(X)$ .*

**Definition 1.25** (Iterated bracketing). *As in  $\text{Br}(X)$  (see Definition 1.14), for  $t_1, t_2 \in \mathcal{L}(X)$ , we will use the notation  $\text{ad}_{t_1}(t_2) := [t_1, t_2]$  for the left bracketing by  $t_1$  and  $\underline{\text{ad}}_{t_1}(t_2) := [t_2, t_1]$  for the right bracketing by  $t_1$ , which allow us to write compactly iterated brackets. For each  $t \in \mathcal{L}(X)$ , the maps  $\text{ad}_t(\cdot)$  and  $\underline{\text{ad}}_t(\cdot)$  are derivations on  $\mathcal{L}(X)$ , so we can use Leibniz’ formula.*

## 1.4 Hall sets and bases

This paragraph introduces Hall sets, the associated Hall bases and some of their properties. Hall bases are of paramount importance, notably due to the nice rewriting algorithm (recalled in Section 2.1) and to their deep link with Lazard’s elimination process, itself linked with the resolution of formal linear differential equations (as illustrated by [42]). They are of course not the only bases of  $\mathcal{L}(X)$  (see Appendix E for a short discussion in this direction).

### 1.4.1 Definition

There are different conventions in the literature for the definition of Hall sets. Indeed, one may decide to swap left and right factors, or swap the order, or both. We follow Viennot’s convention of [48] (also used in control theory by Sussmann [42]), which is different from the original one by Marshall Hall [20] or from Reutenauer’s in [33].

**Definition 1.26.** A Hall set is a subset  $\mathcal{B}$  of  $\text{Br}(X)$  endowed with a total order  $<$  such that

- $X \subset \mathcal{B}$ ,
- for all  $b_1, b_2 \in \text{Br}(X)$ ,  $(b_1, b_2) \in \mathcal{B}$  iff  $b_1, b_2 \in \mathcal{B}$ ,  $b_1 < b_2$  and either  $b_2 \in X$  or  $\lambda(b_2) \leq b_1$ ,
- for all  $b_1, b_2 \in \mathcal{B}$  such that  $(b_1, b_2) \in \mathcal{B}$  then  $b_1 < (b_1, b_2)$ .

**Remark 1.27.** (See [37, §5]) All Hall sets can be built by induction on the length. One starts with the set  $X$  as well as an order on it. To find all Hall elements of length  $n$  given those of smaller length, one adds first all  $(a, b)$  with  $a \in \mathcal{B}$ ,  $|a| = n - 1$ ,  $b \in X$  and  $a < b$ . Then for each bracket  $b = (b_1, b_2) \in \mathcal{B}$  of length  $2 \leq |b| < n$  one adds all the  $(a, b)$  with  $a \in \mathcal{B}$  with  $|a| = n - |b|$  and  $b_1 \leq a < b$ . Finally, one inserts the newly generated elements of length  $n$  into an ordering, maintaining the condition that  $a < (a, b)$ .

**Remark 1.28.** When  $c = (a, (b_1, b_2)) \in \mathcal{B}$  then  $a$  is “sandwiched” between  $b_1$  and  $c$ :  $b_1 \leq a < c$ . Moreover, iterating the third point of the definition yields  $\min(X) = \min(\mathcal{B})$ .

Definition 1.26 is due to Viennot, who also proved that the third item is necessary. It was also known to Shirshov [39] (albeit with an opposite convention). It unifies multiple previous disjoint constructions and narrower definitions, such as the Lyndon basis (see Section 7) and the historical length-compatible Hall sets (see Section 6). We refer the interested reader to [29] and [48] for short accounts of the history of Hall sets.

The importance of this unified definition is linked with the following result.

**Theorem 1.29** (Viennot, [48]). Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. Then  $\mathbb{E}(\mathcal{B})$  is a basis of  $\mathcal{L}(X)$ .

**Example 1.30.** For instance, with  $X = \{X_0, X_1\}$ , the elements of length at most 4 of each Hall set  $\mathcal{B} \subset \text{Br}(X)$  with an order  $<$  such that  $X_0 < X_1$  and  $X < \text{Br}(X) \setminus X$  are:  $X_0, X_1, (X_0, X_1), (X_0, (X_0, X_1)), (X_1, (X_0, X_1)), (X_0, (X_0, (X_0, X_1))), (X_1, (X_0, (X_0, X_1))), (X_1, (X_1, (X_0, X_1)))$ . But  $b := (X_0, (X_1, (X_0, X_1)))$  does not lie in  $\mathcal{B}$  because  $\lambda(\mu(b)) = X_1 > X_0$ . In fact, the following equality holds in  $\mathcal{L}(X)$ :

$$[X_0, [X_1, [X_0, X_1]]] = [[X_0, X_1], [X_0, X_1]] + [X_1, [X_0, [X_0, X_1]]] = [X_1, [X_0, [X_0, X_1]]]. \quad (1.14)$$

This illustrates how Definition 1.26 prevents elements from  $\text{Br}(X)$ , whose evaluations in  $\mathcal{L}(X)$  are linked by Jacobi relations, to appear simultaneously in  $\mathcal{B}$ .



### 1.4.2 Extension of a Hall order

Constructing a Hall set relies on the construction of an order. In the sequel, we will often need to construct Hall sets from appropriate totally ordered subsets of  $\text{Br}(X)$ . We therefore introduce the following notions and extension result.

**Definition 1.31** ( $\lambda$ -stability). *We say that  $A \subset \text{Br}(X)$  is  $\lambda$ -stable when  $\lambda(A \setminus X) \subset A$ .*

**Definition 1.32** (Hall order). *Let  $A$  be a  $\lambda$ -stable subset of  $\text{Br}(X)$ . We say that an order  $<_A$  on  $A$  is a Hall order when  $<_A$  is a total order on  $A$  such that, for every  $t \in A \setminus X$ ,  $\lambda(t) <_A t$ .*

**Proposition 1.33.** *Let  $A$  be a  $\lambda$ -stable non empty subset of  $\text{Br}(X)$  and  $<_A$  a Hall order on  $A$ . There exists a Hall order  $<$  on  $\text{Br}(X)$  that extends  $<_A$ , i.e. for every  $t_1, t_2 \in A$ ,  $t_1 < t_2$  iff  $t_1 <_A t_2$ .*

*Proof.* We consider the set  $\mathcal{A}$  of the Hall-ordered  $\lambda$ -stable subsets  $(B, <_B)$  with  $B \subset \text{Br}(X)$  such that,  $A \subset B$  and for every  $t_1, t_2 \in A$ ,  $t_1 <_B t_2$  iff  $t_1 <_A t_2$ . The set  $\mathcal{A}$  is equipped with the (partial) order  $\ll$  defined by  $B \ll B'$  if  $B \subset B'$  and for every  $t_1, t_2 \in B$ ,  $t_1 <_B t_2$  iff  $t_1 <_{B'} t_2$ .

*Step 1:* We prove that  $\ll$  is an inductive order on  $\mathcal{A}$ . Let  $(B_i)_{i \in I}$  be a family of  $\mathcal{A}$  totally ordered for  $\ll$ . Let  $\tilde{B} := \cup_{i \in I} B_i$ . For  $t_1, t_2 \in \tilde{B}$ , there exists  $i \in I$  such that both  $t_1, t_2 \in B_i$ , then we say that  $t_1 <_{\tilde{B}} t_2$  if  $t_1 <_{B_i} t_2$ . Then  $(\tilde{B}, <_{\tilde{B}})$  belongs to  $\mathcal{A}$  and  $B_i \ll \tilde{B}$  for every  $i \in I$ .

Thus, by Zorn's lemma, one may consider a maximal element  $(B, <)$  in  $\mathcal{A}$ .

*Step 2:* We prove that  $B = \text{Br}(X)$ . By contradiction, assume the existence of  $t \in \text{Br}(X) \setminus B$ . Let  $p_0 := \max\{k \in \mathbb{N}; |\lambda^k(t)| \geq 1\}$  and  $p := \max\{k \in \llbracket 0, p_0 \rrbracket; \lambda^k(t) \notin B\}$ . Then  $\lambda^k(t) \notin B$  for  $k = 0, \dots, p$  and  $\lambda^k(t) \in B$  for  $k = p+1, \dots, p_0$ , because  $B$  is  $\lambda$ -stable. Let  $\bar{B} := B \cup \{\lambda^k(t); k = 0, \dots, p\}$  be equipped with the order  $<$  that extends the order  $<_B$  on  $B$  and satisfies  $\tau < \lambda^p(t) < \dots < \lambda(t) < t$  for every  $\tau \in B$ . Then  $(\bar{B}, <)$   $\in \mathcal{A}$  and  $B \ll \bar{B}$ , which is a contradiction.  $\square$

**Remark 1.34.** *In particular, for every Hall set  $(\mathcal{B}, <)$ , since  $\mathcal{B}$  is  $\lambda$ -stable, the order  $<$ , defined on  $\mathcal{B}$ , can be extended into a Hall order on  $\text{Br}(X)$ . This means that, although all that is required to construct a Hall set  $\mathcal{B}$  is an order on  $\mathcal{B}$  itself, all Hall sets can be seen as being constructed from a Hall order on the whole  $\text{Br}(X)$ . This proves a verification left to the reader in [32, page 85].*

**Remark 1.35.** *Using Zorn's lemma is quite natural in this context, however, it is not necessary as one can also perform the construction by hand. Let  $\prec$  be a given Hall order on  $\text{Br}(X)$ , possibly unrelated with  $<_A$ . Then, one defines an order  $<$  on  $\text{Br}(X)$  by saying that, for  $a, b \in \text{Br}(X)$ ,*

- if  $a, b \in A$ ,  $a < b$  iff  $a <_A b$ ,
- if  $a \in A$  and  $b \in \text{Br}(X) \setminus A$ ,  $a < b$ ,
- if  $a, b \in \text{Br}(X) \setminus A$ ,  $a < b$  iff  $a \prec b$ .

*One easily checks that  $<$  is indeed a Hall order on  $\text{Br}(X)$  using the assumption that  $A$  is  $\lambda$ -stable.*

*This construction requires the a priori knowledge of a Hall order  $\prec$  on  $\text{Br}(X)$ . Such an order obviously exists. Given a total order  $\prec_X$  on  $X$ , one can extend it to  $\text{Br}(X) \setminus X$  using the lexicographic order (see Remark 1.36) on the triple  $(|t|, \lambda(t), \mu(t))$ , which is a Hall order because  $|\lambda(t)| < |t|$  for each  $t \in \text{Br}(X) \setminus X$ .*

*It remains to construct a total order  $\prec_X$  on  $X$ . If  $X$  is countable, by definition there exists an injection  $f : X \rightarrow \mathbb{N}$  and one can define that  $x \prec_X x'$  iff  $f(x) < f(x')$ . If  $X$  is not countable, the existence of a total order on  $X$  is given by the ordering principle, which is strictly weaker than the axiom of choice (see [17]).*

**Remark 1.36.** *In Remark 1.35 and in the sequel, we often construct Hall orders by partitioning the set to be ordered and/or chaining a finite number of preorders. For example, given an order  $\prec_X$  on  $X$ , we define the associated "lexicographic order" on the triple  $(|t|, \lambda(t), \mu(t))$  by  $t_1 < t_2$  iff*

- either  $|t_1| < |t_2|$ ,
- or  $|t_1| = |t_2| = 1$  and  $t_1 \prec_X t_2$ ,
- or  $|t_1| = |t_2| > 1$  and  $\lambda(t_1) < \lambda(t_2)$ ,
- or  $|t_1| = |t_2| > 1$  and  $\lambda(t_1) = \lambda(t_2)$  and  $\mu(t_1) < \mu(t_2)$ .

Such constructions obviously yield reflexive transitive relations, which are moreover total orders provided that enough preorders are combined. Although the definition uses recursively the order on  $\lambda(t)$  and  $\mu(t)$ , this recursion makes sense since one can construct the order by induction on the length of the brackets. Therefore, in the sequel, when handling such orders of lexicographic nature, we mostly focus on checking that they satisfy the Hall condition  $\lambda(t) < t$ .

We conclude this section by an elementary remark on the possibility to construct Hall sets from Hall orders defined only on parts of  $\text{Br}(X)$ .

**Lemma 1.37.** *Let  $G$  be a  $\lambda$ -stable subset of  $\text{Br}(X)$  with  $X \subset G$ , endowed with a (total) Hall order such that, for every  $b_1 < b_2 \in G$ ,  $(b_1, b_2) \in G$ . There exists a unique Hall set  $\mathcal{B} \subset G$  associated with the order on  $G$ .*

*Proof.* By induction on  $\ell \in \mathbb{N}^*$ , we construct and determine the sets  $\mathcal{B}_\ell \subset G$  of elements of  $\mathcal{B}$  with length  $\ell$ . For  $\ell = 1$ ,  $\mathcal{B}_1 = X \subset G$ . Let  $\ell \geq 2$ . We assume  $\mathcal{B}_j$  constructed for every  $j < \ell$ . Let  $\mathcal{B}_\ell := \cup_{1 \leq j \leq \ell-1} \mathcal{B}_\ell^j$  where  $\mathcal{B}_\ell^1 := \{(b_1, b_2); b_1 \in \mathcal{B}_{\ell-1}, b_2 \in \mathcal{B}_1, b_1 < b_2\}$  and  $\mathcal{B}_\ell^j := \{(b_1, b_2); b_1 \in \mathcal{B}_{\ell-j}, b_2 \in \mathcal{B}_j, \lambda(b_2) \leq b_1 < b_2\}$  for  $j = 2, \dots, \ell-1$ . By the last property of  $G$ ,  $\mathcal{B}_\ell \subset G$ .  $\square$

### 1.4.3 Structure constants

Structure constants of an algebra relative to a basis are the expression of the product of pairs of elements of this basis as linear combinations over it. They are for example often used in physics to describe some specific algebras. We recall this classical notion in the context of Hall bases of free Lie algebras.

**Definition 1.38** (Structure constants). *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. Using the convention of Remark 1.24, for any  $a, b \in \mathcal{B}$ , since  $[a, b] \in \mathcal{L}(X)$ , it can be written as a finite linear combination over  $\mathcal{B}$ , say*

$$[a, b] = \sum_{c \in \mathcal{B}} \gamma_{a,b}^c E(c), \quad (1.15)$$

where the coefficients  $\gamma_{a,b}^c \in \mathbb{K}$  and only a finite number of them are non-zero. The set of all coefficients  $\gamma_{a,b}^c$  are called the structure constants of  $\mathcal{L}(X)$  relative to the basis  $\mathcal{B}$ .

Our choice is to measure the size of the structure constants with the help of the following  $\ell^1$  norm of the coordinates of decompositions in the basis.

**Definition 1.39** (Norm relative to a basis). *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. Any  $a \in \mathcal{L}(X)$  can be written as a finite linear combination of basis elements, say*

$$a = \sum_{c \in \mathcal{B}} \alpha_a^c E(c), \quad (1.16)$$

where the coefficients  $\alpha_a^c \in \mathbb{K}$  and only a finite number of them are non-zero. We endow  $\mathcal{L}(X)$  with the following norm

$$\|a\|_{\mathcal{B}} := \sum_{c \in \mathcal{B}} |\alpha_a^c|. \quad (1.17)$$

In particular,

$$\|[a, b]\|_{\mathcal{B}} = \sum_{c \in \mathcal{B}} |\gamma_{a,b}^c|, \quad (1.18)$$

which justifies why the statements of Section 1.2 measure the growth of the structure constants.

**Definition 1.40** (Support). *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. For  $a \in \mathcal{L}(X)$ , we will denote by  $\text{supp } a$  the set of  $c \in \mathcal{B}$  involved in (1.16) such that  $\alpha_a^c \neq 0$ . In particular, for  $a, b \in \mathcal{B}$ ,  $\text{supp}[a, b]$  denotes the set of  $c \in \mathcal{B}$  involved in (1.15) such that  $\gamma_{a,b}^c \neq 0$ .*

#### 1.4.4 Subsets of a Hall set

In the sequel, we will use the following elementary results on subsets of Hall sets.

**Lemma 1.41.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set,  $A$  be a finite subset of  $\mathcal{B}$ . Then, for every  $t \in \mathcal{B} \cap \text{Br}_A$ , we have  $t \geq \min(A)$ .*

*Proof.* The proof is by induction on  $\ell(t) := \min\{|\tau|_A; \tau \in \text{Br}(A), \mathfrak{I}(\tau) = t\} \in \mathbb{N}^*$ . If  $t \in \mathcal{B} \cap \text{Br}_A$  and  $\ell(t) = 1$ , then  $t \in A$  thus the conclusion holds. Now, let  $t \in \mathcal{B} \cap \text{Br}_A$  such that  $\ell(t) \geq 2$ . There exists  $\tau_1, \tau_2 \in \text{Br}(A)$  such that  $t = \mathfrak{I}(\langle \tau_1, \tau_2 \rangle) = (\mathfrak{I}(\tau_1), \mathfrak{I}(\tau_2))$  and  $|\tau_1|_A + |\tau_2|_A = \ell(t)$ . Then  $\mathfrak{I}(\tau_1) \in \mathcal{B} \cap \text{Br}_A$  and  $\ell(\mathfrak{I}(\tau_1)) \leq |\tau_1|_A < \ell(t)$  thus, by the induction assumption,  $\mathfrak{I}(\tau_1) \geq \min(A)$ . Finally,  $t > \mathfrak{I}(\tau_1) \geq \min(A)$ .  $\square$

**Lemma 1.42.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set,  $A \subset \mathcal{B}$  free and  $\Delta$  a finite subset of  $\mathcal{B} \cap \text{Br}_A$  such that, for every  $t \in \Delta \setminus A$ ,  $\lambda(t) < \min(\Delta)$ . Then  $\Delta$  is free.*

*Proof.* If  $t \in A$ , then  $t \notin (\text{Br}_A, \text{Br}_A)$  because  $A$  is free, thus  $t \notin (\text{Br}_\Delta, \text{Br}_\Delta)$  because  $\text{Br}_\Delta \subset \text{Br}_A$ . We assume the existence of  $t \in \Delta \cap (\text{Br}_\Delta, \text{Br}_\Delta)$ . Then  $t \in \Delta \setminus A$  and  $t \in \mathcal{B}$  thus  $\lambda(t) \in \mathcal{B} \cap \text{Br}_\Delta$ . By Lemma 1.41,  $\lambda(t) \geq \min(\Delta)$ , which is a contradiction.  $\square$

## 2 A rough bound for the decomposition algorithm

The main goal of this section is to prove Theorem 1.1, which provides a first rough bound of structure constants of  $\mathcal{L}(X)$  relative to Hall sets. We start by an introduction to the classical decomposition algorithm over Hall sets in Section 2.1 then proceed to the proof of our first estimate in Section 2.2.

### 2.1 The classical decomposition algorithm

Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. For every  $a < b \in \mathcal{B}$ , the bracket  $[a, b]$  can be decomposed in  $\mathcal{L}(X)$  on  $\mathcal{B}$  using the following recursive algorithm:

- if  $b \in X$  or  $\lambda(b) \leq a$ , then  $(a, b) \in \mathcal{B}$  so we are done,
- otherwise, writing  $b = (\lambda(b), \mu(b))$  with  $a < \lambda(b) < \mu(b)$ , the Jacobi identity yields

$$[a, b] = [[a, \lambda(b)], \mu(b)] + [\lambda(b), [a, \mu(b)]], \quad (2.1)$$

then,

- apply the algorithm to decompose  $[a, \lambda(b)]$  as a  $\sum \alpha_d E(d)$  over a finite subset of  $\mathcal{B}$  and, for each  $d$ , apply the algorithm to decompose  $[d, \mu(b)]$  or  $-[\mu(b), d]$  on the basis (depending on whether  $d < \mu(b)$  or  $\mu(b) < d$ ),
- apply the algorithm to decompose  $[a, \mu(b)]$  as a  $\sum_d \alpha_d E(d)$  over a finite subset of  $\mathcal{B}$  and, for each  $d$ , apply the algorithm to decompose  $[\lambda(b), d]$  or  $-[d, \lambda(b)]$  on the basis (depending on whether  $\lambda(b) < d$  or  $d < \lambda(b)$ ),

– return the sum of all these decompositions.

It is not immediate to verify that the recursion terminates. One could hope to make a proof by induction on  $|a| + |b|$ . Indeed, when decomposing  $[a, \lambda(b)]$ , one has  $|a| + |\lambda(b)| < |a| + |b|$ ; but this argument is insufficient since, when decomposing  $[d, \mu(b)]$ , one has  $|d| + |\mu(b)| = |a| + |b|$ .

Nevertheless, it is well-known that this algorithm does converge, thanks to the properties of Hall sets, see e.g. [33, Section 9] for a nice exposition (albeit with reversed order conventions), or [25, 30, 32, 35] for earlier occurrences of this algorithm. The key point is to prove along the induction that, when  $(a, b) \notin \mathcal{B}$ , the elements  $c \in \text{supp}[a, b]$  satisfy  $\lambda(c) > a$ . This allows for an induction on the couple  $(|a| + |b|, \min\{a, b\})$ .

In Section 4.5, we give a cleaner interpretation of the induction process at stake.

In particular, this algorithm yields a constructive proof that  $\mathbb{E}(\mathcal{B})$  indeed spans  $\mathcal{L}(X)$ .

This algorithm is described in many theoretical works (see all references above) and implemented in most Lie algebraic packages (see e.g. `Hall._rewrite_bracket` in SageMath [44], `lie_basis._prod` in CoRoPa's LibAlgebra [5] or `phbize` in LTP [46]). However, its complexity has, up to our knowledge, not been investigated before, even in papers such as [16] where the authors strove to optimize the complexity of the generation of the basis, but without considering the complexity of the decomposition algorithm.

The interpretation of Section 4.5 bounds the size of the associated call stack. The estimates on the structure constants proved in the next paragraph (and in the sequel of the paper) can be seen as being first steps towards estimating the computational complexity of this algorithm (both with respect to space and time). Indeed, assuming that representing in memory a bracket  $b \in \text{Br}(X)$  requires a space of size  $\mathcal{O}(|b|)$ , representing the result of the algorithm, i.e. the decomposition of  $[a, b]$  on the basis, requires a space of size  $\mathcal{O}((|a| + |b|)\|[a, b]\|_{\mathcal{B}})$ , if the algorithm returns a list of (signed) brackets whose sum equals  $[a, b]$  in  $\mathcal{L}(X)$ . Giving a precise estimate of the time-complexity of the decomposition algorithm is more difficult, since it depends on the time-complexity of the comparison operation, which depends on the specific order defining  $\mathcal{B}$ . However, we expect that, given the time-complexity of the comparison operation and the implementation details of the algorithm, adapting the methods we develop in the sequel could lead to bounds on the asymptotic time-complexity of the algorithm.

## 2.2 A first naive bound for structure constants

Even if  $X$  is infinite, any given bracket  $[a, b]$  only involves at most  $|a| + |b|$  elements of  $X$ . Thus, Theorem 1.1 of the introduction is a direct consequence of the following estimate. This estimate is by no means optimal and we will prove much tighter upper bounds in the next sections. Nevertheless, we believe that it is interesting to remark that naively tracking the double induction on  $(|a| + |b|, \min\{a, b\})$  of this classical algorithm does not yield a very practical estimate.

**Theorem 2.1.** *Assume that  $X$  is finite. Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. For every  $a < b \in \mathcal{B}$ ,*

- either  $(a, b) \in \mathcal{B}$  and then  $\|[a, b]\|_{\mathcal{B}} = 1$ ,
- or  $[a, b] = \sum \alpha_c \mathbb{E}(c)$  where the sum is finite,  $\alpha_c \in \mathbb{Z}$ ,  $c \in \mathcal{B} \setminus X$ ,

$$a < \lambda(c) \tag{2.2}$$

and one has the size estimate

$$\|[a, b]\|_{\mathcal{B}} = \sum |\alpha_c| \leq 2^{|X|} \frac{n(n+1)}{2}, \tag{2.3}$$

where  $n = |a| + |b|$ .

*Proof.* The proof consists in applying the decomposition algorithm described in Section 2.1, while keeping track of the number of required iterations. Heuristically, each iteration consists in applying the Jacobi identity, which doubles the number of terms. We construct by induction on  $n \geq 2$  a non-decreasing sequence  $C_n \in [1, +\infty)$  such that the following property holds.

$\mathcal{H}_n$ : “for every  $a < b \in \mathcal{B}$  such that  $|a| + |b| = n$ , either  $(a, b) \in \mathcal{B} \setminus X$ , or  $[a, b] = \sum \alpha_c \mathbb{E}(c)$  where the sum is finite,  $\alpha_c \in \mathbb{Z}$ ,  $c \in \mathcal{B}$ ,  $\lambda(c) > a$  and, in both cases  $\|[a, b]\|_{\mathcal{B}} \leq C_n$ ”.

**Initialization for  $n = 2$ .** Let  $a < b \in \mathcal{B}$  with  $|a| + |b| = 2$ . Then  $a, b \in X$  and  $(a, b) \in \mathcal{B}$  so the conclusion holds with  $C_2 := 1$ .

**Induction on  $n \geq 3$ .** The set

$$\mathcal{G}_n := \{a \in \mathcal{B}; \exists b \in \mathcal{B} \text{ such that } a < b \text{ and } |a| + |b| = n\} \quad (2.4)$$

is finite and  $r := |\mathcal{G}_n| \leq |\{a \in \mathcal{B}; |a| \leq n-1\}| \leq |X|^n$  (for a more precise estimate, albeit generically of the same magnitude, one could use Witt’s formula [50]). Let  $a^1 > \dots > a^r$  be the elements of  $\mathcal{G}_n$  in decreasing order. We prove by induction on  $j \in \llbracket 1, r \rrbracket$  the following property.

$\mathcal{H}'_n(j)$ : “for every  $b \in \mathcal{B}$  such that  $a^j < b$  and  $|a^j| + |b| = n$  then either  $(a^j, b) \in \mathcal{B}$ , or  $[a^j, b] = \sum \alpha_c \mathbb{E}(c)$  where the sum is finite,  $\alpha_c \in \mathbb{Z}$ ,  $c \in \mathcal{B} \setminus X$ ,  $\lambda(c) > a^j$  and, in both cases  $\|[a^j, b]\|_{\mathcal{B}} \leq C'_n(j)$  where  $C'_n(j) := (2C_{n-1})^{j-1}$ ”.

*Initialization for  $j = 1$ .* Let  $b \in \mathcal{B}$  such that  $a^1 < b$  and  $|a^1| + |b| = n$ . If  $b \in X$ , then  $(a^1, b) \in \mathcal{B}$  and the conclusion holds. From now on, we assume that  $b \notin X$ . Then

$$b = (\lambda(b), \mu(b)) \text{ where } \lambda(b) < \mu(b) \in \mathcal{B} \text{ and } \lambda(\mu(b)) \leq \lambda(b) < b. \quad (2.5)$$

If  $\lambda(b) \leq a^1$ , then  $(a^1, b) = (a^1, (\lambda(b), \mu(b))) \in \mathcal{B}$  and the conclusion holds. Now, let us prove that, by definition of  $a^1$ , the situation  $a^1 < \lambda(b)$  cannot happen. Working by contradiction, we assume that  $a^1 < \lambda(b)$ . By Jacobi’s identity

$$[a^1, b] = [a^1, [\lambda(b), \mu(b)]] = [[a^1, \lambda(b)], \mu(b)] + [\lambda(b), [a^1, \mu(b)]]. \quad (2.6)$$

Thus, at least one of the two terms in the right-hand side is nonzero. Let us assume it is the first one. By the induction assumption on  $n$ , we have  $[a^1, \lambda(b)] = \sum \alpha_d \mathbb{E}(d)$  where the sum is finite and non trivial,  $\alpha_d \in \mathbb{Z}$ ,  $d \in \mathcal{B}$  and  $d > \lambda(d) \geq a^1$ . Therefore  $[[a^1, \lambda(b)], \mu(b)] = \sum \alpha_d [d, \mu(b)]$  where the sum is finite and non trivial. Let  $d \in \mathcal{B} \setminus \{\mu(b)\}$  such that  $\alpha_d \neq 0$  and  $d \neq \mu(b)$ . Then  $d$  and  $\mu(b)$  are two distinct elements of  $\mathcal{B}$ , with total length  $n$ , and strictly greater than  $a^1$ , which is in contradiction with the definition of  $a^1$ .

This concludes the proof for  $j = 1$  with  $C'_n(1) := 1 = (2C_{n-1})^0$ .

*Induction on  $j$ .* Let  $j \in \llbracket 2, r \rrbracket$ . We assume that  $\mathcal{H}'_n(k)$  holds for  $k \in \llbracket 1, j-1 \rrbracket$ . Let  $b \in \mathcal{B}$  such that  $a^j < b$  and  $|a^j| + |b| = n$ . If  $b \in X$ , then  $(a^j, b) \in \mathcal{B}$  and the conclusion holds. From now on, we assume that  $b \notin X$ . Then (2.5) holds. If  $\lambda(b) \leq a^j$  then  $(a^j, b) = (a^j, (\lambda(b), \mu(b))) \in \mathcal{B}$  and the conclusion holds. From now on, we assume that  $a^j < \lambda(b)$ . By Jacobi’s identity

$$[a^j, b] = [a^j, [\lambda(b), \mu(b)]] = [[a^j, \lambda(b)], \mu(b)] + [\lambda(b), [a^j, \mu(b)]]. \quad (2.7)$$

Let us decompose both terms.

- *Study of  $[[a^j, \lambda(b)], \mu(b)]$ .* By the induction assumption on  $n$ , we have  $[a^j, \lambda(b)] = \sum \alpha_d \mathbb{E}(d)$  where the sum is finite,  $\alpha_d \in \mathbb{Z}$ ,  $d \in \mathcal{B}$ ,  $\lambda(d) \geq a^j$  and  $\sum |\alpha_d| \leq C_{|a^j| + |\lambda(b)|} \leq C_{n-1}$  (since the sequence is non-decreasing). Then  $[[a^j, \lambda(b)], \mu(b)] = \sum \alpha_d [d, \mu(b)]$ . For every  $d \in \mathcal{B}$  such that  $\alpha_d \neq 0$ ,  $\min\{d, \mu(b)\} > a^j$  (because  $d > \lambda(d) \geq a^j$  and  $\mu(b) > \lambda(b) > a^j$ ), thus it belongs to  $\{a^1, \dots, a^{j-1}\}$ . Therefore  $[d, \mu(b)] = \sum \beta_c^d \mathbb{E}(c)$  where the sum is finite,  $\beta_c^d \in \mathbb{Z}$ ,  $c \in \mathcal{B}$ ,  $\lambda(c) \geq \min\{d, \mu(b)\} > a^j$  (which proves (2.2)) and  $\|[d, \mu(b)]\|_{\mathcal{B}} = \sum |\beta_c^d| \leq C'_n(j-1)$ . (We implicitly use the fact that  $i \mapsto C'_n(i)$  is non-decreasing). Eventually

$$\|[a^j, \lambda(b)], \mu(b)\|_{\mathcal{B}} \leq \sum_{d \in \mathcal{B}} |\alpha_d| \sum_{c \in \mathcal{B}} |\beta_c^d| \leq C_{n-1} C'_n(j-1). \quad (2.8)$$

- *Study of  $[\lambda(b), [a^j, \mu(b)]]$* : Working in the same way,  $\|[\lambda(b), [a^j, \mu(b)]]\|_{\mathcal{B}} \leq C_{n-1}C'_n(j-1)$ .

Finally, we get the expected decomposition of  $[a^j, b]$  together with

$$\| [a^j, b] \|_{\mathcal{B}} \leq 2C_{n-1}C'_n(j-1) = C'_n(j) = (2C_{n-1})^{j-1}. \quad (2.9)$$

Thus,  $\mathcal{H}_n$  holds with

$$C_n := C'_n(r) = (2C_{n-1})^{r-1} \leq (2C_{n-1})^{|X|^{n-1}}. \quad (2.10)$$

**Estimate of  $C_n$ .** Our strategy is to find and estimate a sequence  $D_n$  for  $n \geq 2$  such that  $D_2 \geq C_2$  and  $D_n = (2D_{n-1})^{|X|^{n-1}}$  for  $n \geq 3$ . Assuming temporarily that  $D_{n-1} \geq 2^{|X|^{n-1}}$  (which we check below), the induction relation yields

$$D_n \leq D_{n-1}^{|X|^n}. \quad (2.11)$$

Taking the logarithm twice and solving the nested recurrence yields

$$\ln \ln D_n \leq \ln \ln D_2 + \sum_{k=3}^n k \ln |X|. \quad (2.12)$$

So  $D_n \leq D_2^{|X|^{\frac{n(n+1)}{2}-3}}$ . In particular, choosing  $D_2 = 2^{|X|^{\frac{2(2+1)}{2}}} > 1 = C_2$  yields

$$D_n \leq 2^{|X|^{\frac{n(n+1)}{2}}} \quad (2.13)$$

for every  $n \geq 2$ . Thus we check *a posteriori* that  $D_{n-1} \geq 2^{|X|^{n-1}}$  for every  $n \geq 3$  so the estimate was legitimate.

Then, by construction, one has  $C_n \leq D_n$  for every  $n \geq 2$  which concludes the proof of (2.3).  $\square$

### 3 Iterated brackets over alphabetic subsets

The main goal of this section is to prove Theorem 1.2, which yields a sharp bound for the norm of brackets of elements known to lie in some specific subsets of a Hall set. We start by introducing the notion of *alphabetic subsets* in Section 3.1. Then, we prove a precised version of Theorem 1.2 in Section 3.2. Eventually, we explain in what sense this estimate can be seen as optimal in Section 3.3.

#### 3.1 Alphabetic subsets

**Definition 3.1** (Alphabetic subset). *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set and  $A \subset \mathcal{B}$ . We say that  $A$  is alphabetic when for all  $a_1, a_2 \in A$  such that  $a_1 < a_2$ , we have  $(a_1, a_2) \in \mathcal{B}$ .*

For instance,  $X$  is alphabetic. Every subset of an alphabetic set is also alphabetic.

**Remark 3.2.** *The set of indeterminates  $X$  is sometimes referred to as the “alphabet”. The fact that subsets of  $\mathcal{B}$  satisfying Definition 3.1 behave like  $X$  explains our choice of terminology.*

A characterization of alphabetic subsets of  $\text{Br}(X)$  is given by the following statement.

**Lemma 3.3.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. For every  $a \in \mathcal{B}$ ,  $\mathcal{B}_a := \{a\} \cup \{b \in \mathcal{B}; (a, b) \in \mathcal{B}\}$  is an alphabetic subset of  $\mathcal{B}$ . Conversely, for any finite alphabetic subset  $A$  of  $\mathcal{B}$ , there exists  $a \in \mathcal{B}$  such that  $A \subset \mathcal{B}_a$ .*

*Proof.* Let  $b_1 < b_2 \in \mathcal{B}_a$ . If  $b_1 = a$  or  $b_2 \in X$  then  $(b_1, b_2) \in \mathcal{B}$ . Otherwise  $\lambda(b_2) \leq a < b_1$ , because  $(a, b_2)$  and  $(a, b_1) \in \mathcal{B}$ , thus  $(b_1, b_2) \in \mathcal{B}$ .

Now let  $A \subset \mathcal{B}$  be a finite alphabetic subset. If  $A \subset X$ , let  $a = \min A$ , otherwise let  $a = \max\{\lambda(b); b \in A \setminus X\}$ . The case  $A \subset X$  is straightforward. Assume  $A \not\subset X$ , let  $a^+ \in A \setminus X$  such that  $a = \lambda(a^+)$ . Let  $t \in A$ . We must show that  $t = a$  or  $(a, t) \in \mathcal{B}$ .

- If  $t = a^+$ , then  $a = \lambda(a^+) < a^+ = t$  and  $\lambda(t) = a \leq a$  so  $(a, t) \in \mathcal{B}$ .
- If  $t < a^+$ , since  $A$  is alphabetic,  $(t, a^+) \in \mathcal{B}$  so  $\lambda(a^+) = a \leq t$ . If  $t = a$ , we are done. Else, either  $t \in X$  and thus  $(a, t) \in \mathcal{B}$  or  $t \in A \setminus X$  and, by definition of  $a$ ,  $\lambda(t) \leq a$  so  $(a, t) \in \mathcal{B}$ .
- If  $t > a^+$ , then  $t > a$  because  $a = \lambda(a^+) < a^+$ . Moreover, either  $t \in X$  and thus  $(a, t) \in \mathcal{B}$  or  $t \in A \setminus X$  and, by definition of  $a$ ,  $\lambda(t) \leq a$  so  $(a, t) \in \mathcal{B}$ .

This concludes the proof of the reciprocal statement.  $\square$

**Example 3.4.** If  $X = \{X_0, X_1\}$  and  $\mathcal{B} \subset \text{Br}(X)$  is a Hall set with  $X_0 < X_1$ , then the set  $\{X_0, (X_0, X_1), (X_0, (X_0, X_1))\}$  is an alphabetic subset of  $\mathcal{B}$  because it is a subset of  $\mathcal{B}_{X_0}$ .

The following result (which will be helpful in the sequel) shows that a free alphabetic subset behaves just like a set of indeterminates in the Lie subalgebra that it generates.

**Proposition 3.5.** Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set and  $A \subset \mathcal{B}$  be alphabetic and free. By Lemma 1.19, the canonical surjection  $\mathfrak{I} : \text{Br}(A) \rightarrow \text{Br}_A$  is an isomorphism. Then  $\mathcal{B}' := \mathfrak{I}^{-1}(\mathcal{B} \cap \text{Br}_A)$ , endowed with (the restriction of the preimage by  $\mathfrak{I}$  of) the order on  $\mathcal{B}$ , is a Hall set of  $\text{Br}(A)$ . Moreover, the canonical map  $\mathcal{L}(A) \rightarrow \mathcal{L}(X)$  is an isometry with respect to the norms  $\|\cdot\|_{\mathcal{B}'}$  and  $\|\cdot\|_{\mathcal{B}}$ .

*Proof.* First  $\mathcal{B}'$  is a totally ordered subset of  $\text{Br}(A)$  with the order  $b_1 <' b_2$  iff  $\mathfrak{I}(b_1) < \mathfrak{I}(b_2)$  (in  $\mathcal{B}$ ). Let us check that the three items of Definition 1.26 are satisfied.

- $A \subset \mathcal{B}'$ . Indeed,  $A \subset \mathcal{B}$  and  $\mathfrak{I}$  is the identity on  $A$ .
- For all  $b_1, b_2 \in \text{Br}(A)$ ,  $(b_1, b_2) \in \mathcal{B}'$  iff  $b_1, b_2 \in \mathcal{B}'$ ,  $b_1 <' b_2$  and  $b_2 \in A$  or  $\lambda(b_2) \leq' b_1$ .  
Indeed, let  $b_1, b_2 \in \text{Br}(A)$ . Assume that  $(b_1, b_2) \in \mathcal{B}'$ . Then  $\mathfrak{I}(b_1, b_2) \in \mathcal{B}$ . Since  $\mathcal{B}$  is a Hall set,  $\mathfrak{I}(b_1), \mathfrak{I}(b_2) \in \mathcal{B}$ ,  $\mathfrak{I}(b_1) < \mathfrak{I}(b_2)$  and  $\mathfrak{I}(b_2) \in X$  or  $\lambda(\mathfrak{I}(b_2)) \leq \mathfrak{I}(b_1)$ . Thus,  $b_1, b_2 \in \mathcal{B}'$  and  $b_1 <' b_2$ . Moreover, if  $b_2 \notin A$ , then  $\lambda(b_2) \in \text{Br}(A)$  and  $|b_2|_A > 1$  so  $|\mathfrak{I}(b_2)| > 1$  so  $\mathfrak{I}(b_2) \notin X$  and  $\lambda(\mathfrak{I}(b_2)) \in \mathcal{B}$ . Hence,  $\lambda(\mathfrak{I}(b_2)) \leq \mathfrak{I}(b_1)$  yields  $\lambda(b_2) \leq' b_1$ .  
Conversely, assume that  $b_1, b_2 \in \mathcal{B}'$ ,  $b_1 <' b_2$  and  $b_2 \in A$  or  $\lambda(b_2) \leq' b_1$ . When  $\lambda(b_2) \leq' b_1$ , we obtain that  $(b_1, b_2) \in \mathcal{B}'$  using the fact that  $\mathcal{B}$  is a Hall set. When  $b_2 \in A$ , let  $a := \lambda^k(b_1)$  where  $k \in \mathbb{N}$  is chosen such that  $a \in A$  (iterated left factor, up to falling in  $A$ ). Since  $b_1 \in \mathcal{B}'$ ,  $\mathfrak{I}(b_1) \in \mathcal{B}$  and  $\mathfrak{I}(a) = \lambda^k(\mathfrak{I}(b_1)) \in \mathcal{B}$ . Moreover, by the Hall order property,  $a = \mathfrak{I}(a) \leq \mathfrak{I}(b_1) < \mathfrak{I}(b_2) = b_2$  (with equality if and only if  $k = 0$ ). Since  $A$  is alphabetic,  $(a, b_2) \in \mathcal{B}$ , so  $\lambda(b_2) \leq a$ . Hence  $\lambda(b_2) \leq \mathfrak{I}(b_1)$  and  $\mathfrak{I}(b_1, b_2) \in \mathcal{B}$  so  $(b_1, b_2) \in \mathcal{B}'$ .
- For all  $b_1, b_2 \in \mathcal{B}'$  such that  $(b_1, b_2) \in \mathcal{B}'$ , one has  $b_1 <' (b_1, b_2)$ .  
Indeed, since  $(b_1, b_2) \in \mathcal{B}'$ ,  $\mathfrak{I}(b_1, b_2) \in \mathcal{B}$ , so  $\mathfrak{I}(b_1) < \mathfrak{I}(b_1, b_2)$ .

This shows that  $\mathcal{B}'$  is indeed a Hall set over  $A$ .

By the universal property of the free Lie algebra  $\mathcal{L}(A)$ , there is a canonical morphism of Lie algebras  $h : \mathcal{L}(A) \rightarrow \mathcal{L}(X)$  (which maps any element of  $A$  to itself in  $\mathcal{L}(X)$ ). Since  $\mathcal{B}'$  is a basis of  $\mathcal{L}(A)$  and  $h(\mathcal{B}') \subset \mathcal{B}$  is a linearly independent subset of  $\mathcal{L}(X)$ , this map is injective. By definition, its image is the Lie subalgebra of  $\mathcal{L}(X)$  generated by  $A$ , which is the linear subspace generated by  $\mathcal{B}'$ . Let  $x = \sum_{b \in \mathcal{B}'} \alpha_b E(b)$ , then by definition  $h(x) = \sum_{b \in \mathcal{B}} \alpha_b E(b)$  (where  $\alpha_b = 0$  when  $b \notin \mathcal{B} \cap \text{Br}_A$ ). This shows that this map is indeed an isometry.  $\square$

The following stability property of alphabetic subsets of  $\mathcal{B}$  is a key point of this article.

**Lemma 3.6.** Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set and  $A \subset \mathcal{B}$  be alphabetic. Assume  $a_0 = \min A$  exists, then for all  $a \in A$ ,  $a \neq a_0$ , the set  $A \cup \{(a_0, a)\}$  is alphabetic.

*Proof.* Let  $a \in A$  such that  $a \neq a_0$ . Then since  $A$  is alphabetic,  $(a_0, a) \in \mathcal{B}$ . Hence  $A \cup \{(a_0, a)\} \subset \mathcal{B}$ .

Let  $a_1, a_2 \in A \cup \{(a_0, a)\}$  with  $a_1 < a_2$ . If  $a_1, a_2 \in A$ , then  $(a_1, a_2) \in \mathcal{B}$ , so we may assume that either  $a_1$  or  $a_2$  is equal to  $(a_0, a)$ . If  $a_1 < a_2 = (a_0, a)$ , then the minimality of  $a_0$  implies that  $a_1 \geq a_0$ , so that  $(a_1, a_2) \in \mathcal{B}$ . Now if  $(a_0, a) = a_1 < a_2$ , then since  $A$  is alphabetic, either  $a_2 \in X$  or  $a_2 \notin X$  and  $a_0 \geq \lambda(a_2)$ . In both cases, since  $a_0 < (a_0, a)$ , we have  $(a_1, a_2) \in \mathcal{B}$ .  $\square$

### 3.2 Bound for iterated brackets over alphabetic subsets

**Definition 3.7** (Multisets, i.e. sets with multiplicity). *By convention, we use blackboard bold font to name them and bag brackets to define them, e.g.  $\mathbb{A} := \llbracket 2, 2, 3 \rrbracket$ . Multiplicity matters, so,  $\mathbb{A} \neq \llbracket 2, 3 \rrbracket$ . But order does not matter, so  $\mathbb{A} = \llbracket 2, 3, 2 \rrbracket$ . For such multisets, we define their cardinal as the sum of their multiplicities (so  $|\mathbb{A}| = 3$ ) and their support as the underlying set (so  $\text{supp } \mathbb{A} = \{2, 3\}$ ). If  $\mathbb{A}_1, \mathbb{A}_2$  are two multisets,  $\mathbb{A}_1 + \mathbb{A}_2$  denotes their sum, i.e. the multiset in which the multiplicities are the sum of those within  $\mathbb{A}_1$  and  $\mathbb{A}_2$ . We can also define the difference  $\mathbb{A}_1 - \mathbb{A}_2$  (if the multiplicities in  $\mathbb{A}_1$  are greater than those in  $\mathbb{A}_2$ ). More formally, a multiset whose support is  $A$  is a map  $\mathbb{A} : A \rightarrow \mathbb{N}$ . The cardinal of  $\mathbb{A}$  is  $\sum_{a \in A} \mathbb{A}(a)$ , while the sum (resp. the difference) of the multisets is the sum (resp. the difference, when it exists) of the maps.*

**Definition 3.8** (Trees on a multiset supported in  $\mathcal{B}$ ). *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set and  $\mathbb{A}$  be a multiset with  $A := \text{supp } \mathbb{A} \subset \mathcal{B}$ . We denote by  $\text{Tr}(\mathbb{A})$  the subset of  $\text{Br}(A)$  (and thus of  $\text{Br}(\mathcal{B})$ ) whose elements are brackets of the elements of  $\mathbb{A}$ , involved according to their multiplicity in  $\mathbb{A}$ . Thus, for  $t \in \text{Tr}(\mathbb{A})$ ,  $|t|_A = |t|_{\mathcal{B}} = |\mathbb{A}|$ .*

**Definition 3.9** (Leaves). *For  $A \subset \mathcal{B}$  and  $t \in \text{Br}(A)$ ,  $L_A(t)$  (resp.  $\mathbb{L}_A(t)$ ) denotes the set (resp. multiset) of leaves of  $t$  in  $A$ . In particular  $t \in \text{Tr}(\mathbb{L}_A(t))$  and  $\text{supp } \mathbb{L}_A(t) = L_A(t)$ .*

**Theorem 3.10.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set and  $t \in \text{Br}(\mathcal{B})$  such that  $L_{\mathcal{B}}(t)$  is alphabetic. Then*

$$\|\mathbb{E}(I(t))\|_{\mathcal{B}} \leq (|t|_{\mathcal{B}} - 1)! \quad (3.1)$$

and  $\text{supp } \mathbb{E}(I(t)) \subset I(\text{Tr}(\mathbb{L}_{\mathcal{B}}(t)))$ : the supporting elements are obtained by creating new brackets by combining the leaves (with their multiplicity) of  $t$ .

*Proof.* The proof is by induction on  $n = |t|_{\mathcal{B}}$  and obvious when  $n = 1$ . Now let  $n \geq 2$ , assume that the result holds for all  $k < n$ , and prove that it holds for  $n$ .

Let  $a_0$  be the smallest element of  $L_{\mathcal{B}}(t)$ . There exists  $k \in \{1, \dots, n-1\}$ ,  $a_{i_1}, \dots, a_{i_k} \in \mathbb{L}_{\mathcal{B}}(t)$  and  $v \in \text{Tr}(\llbracket a_{i_1}, \dots, a_{i_k} \rrbracket)$  such that  $\langle a_0, v \rangle$  or  $\langle v, a_0 \rangle$  is a sub-tree of  $t$  and either  $v = a_0$  or  $a_0$  does not appear in  $v$ . In the first case,  $\|\mathbb{E}(I(t))\|_{\mathcal{B}} = 0$ , so we may assume that  $a_0$  does not appear in  $v$ . Using the iterated Jacobi identity, we can write  $[a_0, I(v)] = \sum_{j=1}^k \mathbb{E}(I(v_{i_j}))$ , where  $v_{i_j} \in \text{Br}(\mathcal{B})$  is the bracket obtained from  $v$  by replacing  $a_{i_j}$  by  $\langle a_0, a_{i_j} \rangle$ . Then, by bilinearity of  $[\cdot, \cdot]$ , we get  $\mathbb{E}(I(t)) = \sum_{j=1}^k \mathbb{E}(I(t_j))$ , where for all  $j \in \llbracket 1, k \rrbracket$ ,  $t_j \in \text{Br}(\mathcal{B})$  and  $\mathbb{L}_{\mathcal{B}}(t_j) = \mathbb{L}_{\mathcal{B}}(t) + \llbracket (a_0, a_{i_j}) \rrbracket - \llbracket a_0, a_{i_j} \rrbracket$ . By Lemma 3.6,  $L_{\mathcal{B}}(t_j)$  is alphabetic. by the induction hypothesis, for all  $j \in \llbracket 1, k \rrbracket$ , we have  $\|\mathbb{E}(I(t_j))\|_{\mathcal{B}} \leq (n-2)!$  and  $\text{supp } \mathbb{E}(I(t_j)) \subset I(\text{Tr}(\mathbb{L}_{\mathcal{B}}(t_j))) \subset I(\text{Tr}(\mathbb{L}_{\mathcal{B}}(t)))$ . The expression  $\mathbb{E}(I(t)) = \sum_{j=1}^k \mathbb{E}(I(t_j))$  and the fact that  $k < n$  then show that  $\|\mathbb{E}(I(t))\|_{\mathcal{B}} \leq (n-1)!$ .  $\square$

### 3.3 Optimality case

**Remark 3.11.** *For  $b \in \text{Br}(A)$ ,  $L_A(b)$  denotes the set of leaves in  $A$  of the tree  $b$ . When  $A$  is a free subset, this notion can be transported on  $\text{Br}_A$ , because  $\mathfrak{i} : \text{Br}(A) \rightarrow \text{Br}_A$  is an isomorphism. Setting  $L_A(\mathfrak{i}(b)) := L_A(b)$  for  $b \in \text{Br}(A)$  extends  $L_A$  to  $\text{Br}_A$ .*

**Proposition 3.12.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set,  $n \in \mathbb{N}^*$ ,  $a_1 < \dots < a_n \in \mathcal{B}$ . We assume that*

- $A := \{a_1, \dots, a_n\}$  is a free alphabetic subset of  $\mathcal{B}$ ,
- for all  $b_1, b_2 \in \mathcal{B} \cap \text{Br}_A$  with  $L_A(b_1) \cap L_A(b_2) = \emptyset$  then  $b_1 < b_2$  iff  $\max L_A(b_1) < \max L_A(b_2)$ .

Then

$$\|\llbracket \dots \llbracket a_n, a_{n-1} \rrbracket, \dots, a_1 \rrbracket\|_{\mathcal{B}} = (n-1)!. \quad (3.2)$$



*Proof.* We proceed by induction on  $n \in \mathbb{N}^*$ . The conclusion holds for  $n = 1$ . Let  $n \geq 2$ . We assume the property proved up to  $(n - 1)$ . Let  $A = \{a_1 < \dots < a_n\}$  be a free alphabetic subset of  $\mathcal{B}$  such that the order of  $\mathcal{B}$  is compatible with  $\max L_A$  (in the sense of the second point above). Let  $w \in \text{Br}(A)$  be defined by  $w := \langle \dots \langle a_n, a_{n-1} \rangle, \dots, a_1 \rangle$ .

For any  $k \in \llbracket 2, n \rrbracket$ , we have  $(a_1, a_k) \in \mathcal{B}$  because  $A$  is alphabetic and, by compatibility of the Hall set order  $<$  with  $\max L_A$ ,  $a_{k-1} < (a_1, a_k) < a_{k+1}$ . Thus  $a_2 < \dots < a_{k-1} < (a_1, a_k) < a_{k+1} < \dots < a_n$ . Let  $\mathbb{A}_k := \lceil a_2, \dots, a_{k-1}, (a_1, a_k), a_{k+1}, \dots, a_n \rceil$  and  $A_k = \text{supp}(\mathbb{A}_k)$ . By Lemma 3.6,  $A_k$  is alphabetic. By Lemma 1.42,  $A_k$  is free. Indeed,  $A_k \subset \mathcal{B} \cap \text{Br}_A$  and  $a_1 = \lambda((a_1, a_k)) < \min(A_k)$ .

*Step 1:* We prove the compatibility of the order  $<$  with  $\max L_{A_k}$ . Let  $b_1, b_2 \in \mathcal{B} \cap \text{Br}_{A_k}$  with  $L_{A_k}(b_1) \cap L_{A_k}(b_2) = \emptyset$ . We assume  $\max L_{A_k}(b_1) < \max L_{A_k}(b_2)$ . Let us prove that  $b_1 < b_2$ .

We have  $b_1, b_2 \in \mathcal{B} \cap \text{Br}_A$  and  $L_A(b_1) \cap L_A(b_2) = \emptyset$ . Moreover,  $\max L_A(b_j) = \max L_{A_k}(b_j)$  except when  $\max L_{A_k}(b_j) = (a_1, a_k)$  and then  $\max L_A(b_j) = a_k$ . In any case,  $\max L_A(b_1) < \max L_A(b_2)$  thus  $b_1 < b_2$  because  $(a_1, a_k)$  is inserted at the position of  $a_k$  thanks to the order on  $A$ .

*Step 2:* We apply the induction assumption. Let  $w_k \in \text{Br}(A_k)$  be defined by

$$w_k := \langle \dots \langle \langle \langle a_n, a_{n-1} \rangle, \dots, a_{k+1} \rangle, (a_1, a_k) \rangle, a_{k-1} \rangle \dots, a_2 \rangle. \quad (3.3)$$

Then  $\|\mathbb{E}(I(w_k))\|_{\mathcal{B}} = (n - 2)!$ .

*Step 3:* We prove that the sets  $\mathcal{B} \cap I(\text{Tr}(\mathbb{A}_k))$  for  $k \in \llbracket 2, n \rrbracket$  are two by two disjoint. Let  $j \neq k \in \llbracket 2, n \rrbracket$ ,  $B_j \in \text{Tr}(\mathbb{A}_j)$  and  $B_k \in \text{Tr}(\mathbb{A}_k)$  such that  $b_j = I(B_j)$  and  $b_k = I(B_k)$  belong to  $\mathcal{B}$ . Then  $B_j$  and  $B_k$  are two different elements of  $\text{Br}(\Delta)$  where

$$\Delta := A_j \cup A_k = \{a_2, \dots, a_n, (a_1, a_j), (a_1, a_k)\}. \quad (3.4)$$

By Lemma 1.42,  $\Delta$  is free because  $\Delta \subset \mathcal{B} \cap \text{Br}_A$ ,  $A$  is alphabetic and  $a_1 < \min(\Delta)$ . Thus  $I$  is injective on  $\text{Br}(\Delta)$  and  $b_j \neq b_k$ .

*Step 4: Conclusion.* Iterating Jacobi's identity proves that  $-\mathbb{E}(I(w)) = \mathbb{E}(I(w_2)) + \dots + \mathbb{E}(I(w_n))$ . By Steps 2 and 3,  $\|\mathbb{E}(I(w))\|_{\mathcal{B}} = (n - 1)!$ .  $\square$

**Corollary 3.13.** *Let  $n \geq 2$  and  $X = \{X_1, \dots, X_n\}$ . There exists a Hall set  $\mathcal{B} \subset \text{Br}(X)$  such that*

$$\|[\dots [X_n, X_{n-1}], \dots, X_1]\|_{\mathcal{B}} = (n - 1)!. \quad (3.5)$$

*Proof.* For  $b \in \text{Br}(X)$ , we use the notation  $L(b) \subset X$  to denote the set of leaves involved in  $b$ .

The strategy consists in constructing a Hall order  $<$  on  $\text{Br}(X)$  such that  $X_1 < \dots < X_n$  and, for every  $b_1, b_2 \in \text{Br}(X)$  with  $L(b_1) \cap L(b_2) = \emptyset$ , then  $b_1 < b_2$  iff  $\max L(b_1) < \max L(b_2)$ . Then, considering the Hall set  $\mathcal{B}$  of  $\text{Br}(X)$  associated with this order and applying Proposition 3.12 to the free subset  $A := \{X_1, \dots, X_n\}$  yields (3.5).

Let  $\prec$  be any Hall order on  $\text{Br}(X)$ . Such an order does exist: see the second paragraph of Remark 1.35 or Proposition 1.33. We will use  $\prec$  as an arbitrary order to compare brackets for which we have no particular requirement. For  $b_1 \neq b_2 \in \text{Br}(X)$ , we write  $b_1 < b_2$  when

- either  $b_1 = X_j$  and  $b_2 = X_k$  with  $j < k$  (i.e.  $X_1 < \dots < X_n$ ),
- or  $\max L(b_1) < \max L(b_2)$ ,
- or  $\max L(b_1) = \max L(b_2)$  and  $b_1 \prec b_2$ .

Since  $<$  is defined as the lexicographic order on the couple  $(\max L, \prec)$ , it is a total order on  $\text{Br}(X)$  which satisfies the desired properties.

It remains to check that  $<$  is indeed a Hall order. Let  $t \in \text{Br}(X)$  with  $|t| \geq 2$ . If  $\max L(\lambda(t)) < \max L(t)$  then  $\lambda(t) < t$ . Otherwise,  $\max L(\lambda(t)) = \max L(t)$  and  $\lambda(t) \prec t$  since  $\prec$  is a Hall order, thus  $\lambda(t) < t$ .  $\square$

## 4 A refined bound stemming from brackets structure

The main goal of this section is to prove Theorem 1.3, which provides a sharp bound for the structure constants of  $\mathcal{L}(X)$  relative to Hall sets in the general case. We start by introducing the notion of *relative folding* in Section 4.1 and estimate its length in Section 4.2. Then, we prove a precised version of Theorem 1.3 in Section 4.3. We explain in what sense this estimate can be seen as optimal in Section 4.4. Eventually, we prove in Section 4.5 that the *length of the relative folding* provides a natural strictly decreasing indexation of the classical recursive decomposition algorithm of Section 2.1.

### 4.1 Relative folding

Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. We define a notion of *relative folding* which will be of paramount importance for tracking, during the execution of the decomposition algorithm of Section 2.1, which brackets fall directly in the basis and which must be split in two using Jacobi's identity.

**Definition 4.1.** For  $a < b \in \mathcal{B}$ , the relative folding of  $b$  with respect to  $a$  is the tree  $T_a(b) \in \text{Br}(\mathcal{B})$  defined by induction by

$$T_a(b) := \begin{cases} b & \text{when } (a, b) \in \mathcal{B}, \\ \langle T_a(\lambda(b)), T_a(\mu(b)) \rangle & \text{otherwise,} \end{cases} \quad (4.1)$$

which makes sense as, when  $(a, b) \notin \mathcal{B}$ , then  $b \notin X$  and  $a < \lambda(b) < \mu(b)$ .

**Lemma 4.2.** For each  $c \in L_{\mathcal{B}}(T_a(b))$ ,  $(a, c) \in \mathcal{B}$ . Moreover,  $b = \mathbf{1}(T_a(b))$ .

*Proof.* These are immediate consequences of Definition 4.1.  $\square$

**Proposition 4.3.** For  $a < b \in \mathcal{B}$ ,  $\{a\} \cup L_{\mathcal{B}}(T_a(b))$  is an alphabetic subset of  $\mathcal{B}$ .

*Proof.* By Lemma 4.2, the considered set is a subset of  $\mathcal{B}_a$ , which is alphabetic by Lemma 3.3.  $\square$

**Definition 4.4.** For  $a < b \in \mathcal{B}$ , we define  $\theta_a(b) := |T_a(b)|_{\mathcal{B}} \in \mathbb{N}^*$ .

**Example 4.5.** Assume that  $\mathcal{B} \subset \text{Br}(\{X_0, X_1\})$  is a Hall set such that the tree of Example 1.13,  $t := (((X_0, X_1), ((X_0, X_1), X_1)), ((X_0, X_1), X_1))$  belongs to  $\mathcal{B}$ . Then  $a := (X_0, X_1) \in \mathcal{B}$  and  $T_a(t) = \langle b_1, b_2 \rangle$ , where  $b_1 = ((X_0, X_1), ((X_0, X_1), X_1))$  and  $b_2 = ((X_0, X_1), X_1)$ . So  $\theta_a(t) = 2$ . This is illustrated by the following trees:

$$t = \begin{array}{c} \diagup \quad \diagdown \\ \begin{array}{c} \diagup \quad \diagdown \\ X_0 \quad X_1 \end{array} \quad \begin{array}{c} \diagup \quad \diagdown \\ \begin{array}{c} \diagup \quad \diagdown \\ X_0 \quad X_1 \end{array} \quad X_1 \end{array} \quad \begin{array}{c} \diagup \quad \diagdown \\ X_0 \quad X_1 \end{array} \quad X_1 \end{array} \quad (4.2)$$

and

$$T_a(t) = \begin{array}{c} \diagup \quad \diagdown \\ ((X_0, X_1), ((X_0, X_1), X_1)) \quad ((X_0, X_1), X_1) \end{array} \quad (4.3)$$

## 4.2 Properties of the length of the folding

In order to achieve our goal of proving estimates depending on  $|b|$ , we prove in this paragraph bounds relating  $\theta_a(b)$  and  $|b|$ , along with other elementary remarks on  $\theta$ . These results will be used also in the sequel of the paper for other purposes. We start with the following elementary result.

**Lemma 4.6.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set and  $a, b \in \mathcal{B}$  such that  $(a, b) \in \mathcal{B}$ . There exists a unique  $r = r(a, b) \in \mathbb{N}^* \cup \{+\infty\}$  such that, for each  $n \in \llbracket 0, r \rrbracket$ ,  $\underline{\text{ad}}_b^n(a) \in \mathcal{B}$  and, for each  $n > r$ ,  $\underline{\text{ad}}_b^{n-r} \underline{\text{ad}}_b^r(a) \in \mathcal{B}$ . Moreover, for each  $n \in \llbracket 0, r-1 \rrbracket$ ,  $\underline{\text{ad}}_b^n(a) < b$  and, if  $r$  is finite,  $b < \underline{\text{ad}}_b^r(a)$ .*

*Proof.* Assume that  $\{\underline{\text{ad}}_b^n(a); n \in \mathbb{N}\} \not\subset \mathcal{B}$ . Let  $r$  be the smallest integer such that  $\underline{\text{ad}}_b^{r+1}(a) \notin \mathcal{B}$ . By definition, for all  $n \in \llbracket 0, r \rrbracket$ ,  $\underline{\text{ad}}_b^n(a) \in \mathcal{B}$ , thus  $\underline{\text{ad}}_b^n(a) < b$  for  $n \in \llbracket 0, r-1 \rrbracket$ . Assume  $\underline{\text{ad}}_b^r(a) < b$ , then by hypothesis  $b \notin X$  (otherwise  $\underline{\text{ad}}_b^{r+1}(a) \in \mathcal{B}$ ), and since  $\underline{\text{ad}}_b^r(a) > \lambda(\underline{\text{ad}}_b^r(a)) = \underline{\text{ad}}_b^{r-1}(a) \geq \lambda(b)$  because  $\underline{\text{ad}}_b^r(a) \in \mathcal{B}$ , one sees that  $(\underline{\text{ad}}_b^r(a), b) \in \mathcal{B}$  which is a contradiction. Hence,  $b < \underline{\text{ad}}_b^r(a)$ , and  $b > \underline{\text{ad}}_b^{r-1}(a) = \lambda(\underline{\text{ad}}_b^r(a))$ , so that  $(b, \underline{\text{ad}}_b^r(a)) \in \mathcal{B}$ , and  $\text{ad}_b^p(\underline{\text{ad}}_b^r(a)) \in \mathcal{B}$  for all  $p \in \mathbb{N}$ .  $\square$

**Proposition 4.7.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set.*

1. *If  $|X| \geq 3$  then, for every  $a < b \in \mathcal{B}$ ,  $\theta_a(b) \leq |b|$  and equality holds for arbitrarily long  $b \in \mathcal{B}$ .*
2. *If  $|X| = 2$  then, for every  $a < b \in \mathcal{B}$  with  $|b| \geq 2$ ,  $\theta_a(b) \leq |b| - 1$  and equality holds for arbitrarily long  $b \in \mathcal{B}$ .*

*Proof.* Let us prove the items successively.

1. By Lemma Lemma 4.2,  $|b|$  is the sum of the lengths of the leaves of  $T_a(b)$ , that are in number  $\theta_a(b)$  and each leaf has length at least one 1, one must have  $|b| \geq \theta_a(b)$ .

If  $X$  has at least 3 elements  $X_1 < X_2 < X_3$  then,  $\text{ad}_{X_2}^p(X_3) \in \mathcal{B}$  for every  $p \in \mathbb{N}$ . We prove by induction on  $p \in \mathbb{N}$  that  $\theta_{X_1}(\text{ad}_{X_2}^p(X_3)) = p + 1$ . For  $p = 0$ ,  $\theta_{X_1}(X_3) = 1$  because  $(X_1, X_3) \in \mathcal{B}$ . For  $p \geq 2$ ,  $\theta_{X_1}(\text{ad}_{X_2}^p(X_3)) = \theta_{X_1}(X_2) + \theta_{X_1}(\text{ad}_{X_2}^{p-1}(X_3)) = 1 + p$ .

2. We assume  $X = \{X_0, X_1\}$  with  $X_0 < X_1$ . Then,  $X_0$  is the minimal element of  $\mathcal{B}$ . We prove the estimate by induction on  $|b| \geq 2$ . If  $|b| = 2$  then  $b = (X_0, X_1)$  and  $X_0 \leq a$  thus  $(a, b) \in \mathcal{B}$  and  $\theta_a(b) = 1 = |b| - 1$ . If  $|b| \geq 3$ , then  $|\lambda(b)| \geq 2$  or  $|\mu(b)| \geq 2$ , thus, using the first statement and the induction assumption  $\theta_a(b) = \theta_a(\lambda(b)) + \theta_a(\mu(b)) \leq |\lambda(b)| + |\mu(b)| - 1 = |b| - 1$ .

Let  $b_n$  be the unique element of  $\mathcal{B}$  containing  $X_0$  exactly once and  $X_1$  exactly  $n$  times, whose form is given in Lemma 4.6 and depends on  $r(X_0, X_1)$ . Then  $|b_n| = n + 1$  and we easily get  $\theta_{X_0}(b_n) = n$  by induction on  $n \in \mathbb{N}^*$ .  $\square$

**Lemma 4.8.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. For all  $a \leq \tilde{a} < b \in \mathcal{B}$ , one has  $\theta_{\tilde{a}}(b) \leq \theta_a(b)$ .*

*Proof.* We proceed by induction on  $n := \theta_a(b)$ .

*Initialization for  $n = 1$ .* Either  $b \in X$ , or  $|b| \geq 2$  and  $\lambda(b) \leq a \leq \tilde{a}$ . In both cases,  $\theta_{\tilde{a}}(b) = 1$ .

*Inductive step for  $n \geq 2$ .* Let  $a \leq \tilde{a} < b$  such that  $\theta_a(b) = n$ . If  $\theta_{\tilde{a}}(b) = 1$ , the inequality holds. Otherwise,  $a \leq \tilde{a} < \lambda(b)$ , so  $\theta_a(b) = \theta_a(\lambda(b)) + \theta_a(\mu(b))$  and  $\theta_{\tilde{a}}(b) = \theta_{\tilde{a}}(\lambda(b)) + \theta_{\tilde{a}}(\mu(b))$ , and thus the estimate follows from  $\theta_{\tilde{a}}(\lambda(b)) \leq \theta_a(\lambda(b))$  and  $\theta_{\tilde{a}}(\mu(b)) \leq \theta_a(\mu(b))$ .  $\square$

## 4.3 Bound for a bracket of two basis elements

### 4.3.1 Warm-up version and structure of the supporting basis elements

We start with a warm-up version of our main estimate Theorem 1.3. Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set and  $a < b \in \mathcal{B}$ . By Lemma 4.2,  $[a, b] = \mathbb{E}(I(t))$  where  $t = \langle a, T_a(b) \rangle$  is a bracket of length

$|t|_{\mathcal{B}} = \theta_a(b) + 1$  over  $\{a\} \cup L_{\mathcal{B}}(T_a(b))$  which is an alphabetic subset of  $\mathcal{B}$  by Proposition 4.3. By Theorem 3.10 and the first item of Proposition 4.7,

$$\|[a, b]\|_{\mathcal{B}} \leq \theta_a(b)! \leq |b|!. \quad (4.4)$$

Moreover, Theorem 3.10 also proves that  $\text{supp}[a, b] \subset \mathfrak{I}(\text{Tr}(\mathbb{A}))$  where  $\mathbb{A} := \mathfrak{I}a \mathfrak{I} + \mathbb{L}_{\mathcal{B}}(T_a(b))$ . This is a strong information on the structure of these supporting elements: they are obtained by creating new brackets with a single  $a$  and the leaves (with their multiplicity) of  $T_a(b)$ .

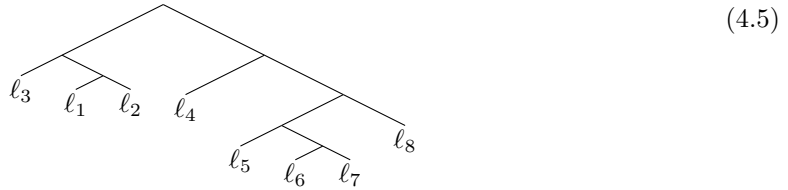
Estimate (4.4) is nice because it is straight-forward and much better than (2.3). However, it is not sharp. This boils down to the fact that, contrary to Theorem 3.10 where no assumption is made on the structure of the considered bracket, here,  $T_a(b)$  models  $b$ , an element of the Hall set  $\mathcal{B}$ . So there is a little more structure in  $\langle a, T_a(b) \rangle$  than the mere fact that its leaves form an alphabetic subset of  $\mathcal{B}$ . We exploit this remark in the next paragraph to tighten the bound from  $\theta_a(b)!$  down to  $\lfloor e(\theta_a(b) - 1)! \rfloor$ .

### 4.3.2 Sharp version

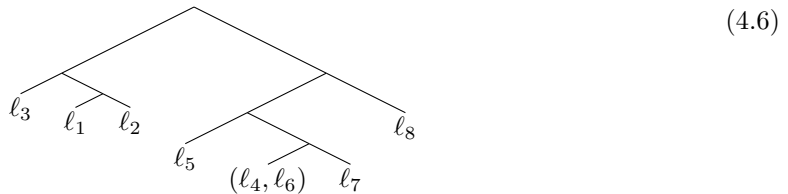
In this paragraph, we prove an enhanced, sharp version of Theorem 1.3. We start by introducing the following operation, linked with the Jacobi identity, which will be helpful for the proof.

**Definition 4.9** (Jacobi distribution). *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. Let  $t \in \text{Br}(\mathcal{B})$  and  $\ell_1, \ell_2 \in \mathcal{B}$  such that  $(\ell_1, \ell_2) \in \mathcal{B}$ , where  $\ell_1$  is a localized leaf of  $t$  and  $\ell_2$  is a localized leaf of the sibling  $t_1$  of  $\ell_1$  in  $t$  (see Remark 4.11). Let  $t_2$  be the tree constructed from  $t_1$  where  $\ell_2$  has been replaced by  $(\ell_1, \ell_2)$ . We denote by  $J(t, \ell_1, \ell_2)$  the tree constructed from  $t$ , where the subtree  $\langle \ell_1, t_1 \rangle$  (or  $\langle t_1, \ell_1 \rangle$ ) has been replaced by  $t_2$ . In particular, one has  $|J(t, \ell_1, \ell_2)|_{\mathcal{B}} = |t|_{\mathcal{B}} - 1$ .*

**Example 4.10.** *Let  $t$  be defined as*



Then  $J(t, \ell_4, \ell_6)$  is the tree



**Remark 4.11.** *In Definition 4.9, “localized leaf” means that the position of the leaf within the tree matters. There will be no ambiguity when using this notation in the sequel concerning which occurrence of  $\ell_1$  and  $\ell_2$  we wish to modify. The “sibling” of a localized leaf  $\ell$  in a tree  $t$  denotes the other subtree sharing the same parent as  $\ell$ .*

*This wording could be formalized using notions coming from graph theory and labeled trees, but we consider that, in our context, a full formalization would make the comprehension harder.*

We can now prove the following estimate, which of course implies Theorem 1.3 since  $\theta_a(b) \leq |b|$  by the first item of Proposition 4.7. In the particular case  $|X| = 2$ , the refined estimate  $\theta_a(b) \leq |b| - 1$  (second item of Proposition 4.7) yields an even smaller upper bound.

**Theorem 4.12.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. For any  $a < b \in \mathcal{B}$ ,*

$$\|[a, b]\|_{\mathcal{B}} \leq \lfloor e(\theta_a(b) - 1)! \rfloor. \quad (4.7)$$

Moreover,  $\text{supp}[a, b] \subset \mathcal{I}(\text{Tr}(\mathbb{A}))$  where  $\mathbb{A} = \langle a \rangle + \mathbb{L}_{\mathcal{B}}(T_a(b))$ .

*Proof.* The part concerning the structure of supporting basis elements has already been proved in Section 4.3.1. So we only need to prove the size estimate.

Let  $n := \theta_a(b)$ . When  $n = 1$ ,  $(a, b) \in \mathcal{B}$  so  $\|[a, b]\|_{\mathcal{B}} = 1$ . When  $n = 2$ ,  $(a, b) = (a, (b_1, b_2))$  where  $b_1, b_2 \in \mathcal{B}$  and the set  $\{a, b_1, b_2\}$  is alphabetic. Hence, by Theorem 3.10,  $\|[a, b]\|_{\mathcal{B}} \leq (3 - 1)! = 2 = \lfloor e! \rfloor$ . We now assume that  $n \geq 3$ .

*Strategy.* The proof is inspired by the method deployed in the proof of Theorem 3.10: seeing  $[a, b]$  as  $\langle a, T_a(b) \rangle$ , a tree of  $\text{Br}(\mathcal{B})$  over an alphabetic subset of  $\mathcal{B}$ , one iteratively selects its minimal leaf, distributes it on the leaves of its sibling using the Jacobi identity, remarks that the leaves of the new tree are an alphabetic subset of  $\mathcal{B}$  and that the new tree is strictly shorter, and continues the process. As announced in Section 4.3.1, this sharper estimate comes from the additional structure of  $T_a(b)$  inherited from the fact that  $b \in \mathcal{B}$ . Thus, the key features of the following argument are to identify this additional structure, prove that it is preserved during the iterative process, and eventually use it to tighten the bound from  $n!$  down to  $\lfloor e(n - 1)! \rfloor$ .

*Structure of  $T_a(b)$ .* Let  $b_1, \dots, b_n \in \mathcal{B}$  be the (not necessarily distinct) leaves of  $T_a(b)$ . By Lemma 4.2, recall that  $b = \mathcal{I}(T_a(b))$ . Let us prove that there exist  $i \neq j \in \llbracket 1, n \rrbracket$  such that  $\langle b_i, b_j \rangle$  is a subtree of  $T_a(b)$  and  $b_i$  is minimal among  $b_1, \dots, b_n$ . Let  $b_k$  be a minimal leaf of maximal depth. If its sibling is also a leaf  $b_{k'}$ , then, since  $b \in \mathcal{B}$ ,  $b_k < b_{k'}$  and the couple  $(i, j) := (k, k')$  is a valid choice. If its left sibling is a tree  $w$ ,  $\langle w, b_k \rangle$  is a subtree of  $T_a(b)$ . But, since  $\mathcal{I}(\langle w, b_k \rangle) \in \mathcal{B}$ ,  $\mathcal{I}(w) < b_k$  so there exists a strictly smaller leaf within  $w$ , which contradicts the minimality of  $b_k$ . If its right sibling is a tree  $w$ ,  $\langle b_k, w \rangle$  is a subtree of  $T_a(b)$ . But, since  $\mathcal{I}(\langle b_k, w \rangle) \in \mathcal{B}$ ,  $\lambda(\mathcal{I}(w)) \leq b_k$  so one can find a leaf smaller or equal to  $b_k$  within  $w$  of greater depth. Eventually, up to reindexing, we can assume that  $T_a(b)$  contains  $\langle b_1, b_2 \rangle$  as a subtree, where  $b_1 \leq b_j$  for every  $j \in \llbracket 1, n \rrbracket$ .

For example, this yields the following structure, with  $b_1$  minimal:

$$T_a(b) = \begin{array}{c} \diagup \quad \diagdown \\ \begin{array}{ccccccc} b_3 & & b_1 & b_2 & & b_4 & \\ & & \diagdown & \diagup & & & \\ & & b_5 & b_6 & b_7 & & b_8 \end{array} \end{array} \quad (4.8)$$

For this tree structure, we knew from the start that  $b_1$  or  $b_6$  was minimal (and we chose  $b_1$  up to re-indexing). Indeed, using repeatedly the axioms of a Hall set and the fact that  $b \in \mathcal{B}$  yields:  $b_1 < b_2$ ,  $b_1 \leq b_3$ ,  $b_6 < b_7$ ,  $b_6 \leq b_5$ ,  $b_5 < (b_5, (b_6, b_7)) < b_8$  and  $(b_5, (b_6, b_7)) \leq b_4$ .

*Paths within  $T_a(b)$ .* Let  $p \in \llbracket 0, n - 1 \rrbracket$ . We say that  $\pi = (\pi_1, \dots, \pi_p) \in \llbracket 2, n \rrbracket^p$  is a path within  $T_a(b)$  when, for each  $i \in \llbracket 1, p - 1 \rrbracket$ , the sibling of  $b_{\pi_i}$  is not a leaf and contains  $b_{\pi_{i+1}}$  as a (deeper) leaf. For  $\pi = (\pi_1, \dots, \pi_p)$  a path of length  $p \geq 1$ , we define  $\pi' = (\pi_1, \dots, \pi_{p-1})$  which is a path of length  $(p - 1)$ .

For the tree given in (4.8), examples of such paths are  $\emptyset$ , (3), (4, 6) or (4, 5, 7). However, (3, 6) is not a path for this definition because  $b_6$  is not a leaf of the sibling of  $b_3$  (which is  $\langle b_1, b_2 \rangle$ ).

There are at most  $\binom{n-1}{p}$  such paths of length  $p$ . Indeed, given a subset of  $p$  elements of  $\llbracket 2, n \rrbracket$ , there is at most a single permutation of its elements such that the property “each next index must be a leaf of the sibling tree” is satisfied.

*Iterative construction.* We intend to define by induction on  $p \in \llbracket 0, n - 1 \rrbracket$ , a set  $\Pi_p$  of admissible paths of length  $p$  and, for each  $\pi \in \Pi_p$ , an element  $a_\pi \in \mathcal{B}$  and a bracket  $B_\pi \in \text{Br}(\mathcal{B})$  such that

- $\mathbb{L}_{\mathcal{B}}(B_\pi) = \wr a_\pi \wr + \wr b_j \wr$ ;  $j \in \llbracket 1, n \rrbracket \setminus \pi$  and in particular  $|B_\pi|_{\mathcal{B}} = n - p + 1$ ,
- $\mathbb{L}_{\mathcal{B}}(B_\pi)$  is alphabetic,
- the right-sibling of  $b_1$  in  $B_\pi$  is a leaf,
- if  $p \geq 1$ , the sibling of  $a_\pi$  in  $B_\pi$  is the sibling of  $b_{\pi_p}$  in  $T_a(b)$ ,
- $\min\{b_1, a_\pi\} = \min \mathbb{L}_{\mathcal{B}}(B_\pi)$ .

For  $p = 0$ , we let  $\Pi_0 := \{\emptyset\}$ ,  $a_\emptyset := a$  and  $B_\emptyset := \langle a, T_a(b) \rangle$  and all properties are satisfied thanks to the remarks made above concerning the structure of  $T_a(b)$ .

Let  $p \in \llbracket 1, n-1 \rrbracket$ . We assume that we have constructed  $\Pi_{p-1}$  and for  $\pi \in \Pi_{p-1}$ ,  $a_\pi$  and  $B_\pi$  with the claimed properties. We now define  $\Pi_p$  as the set of paths  $\pi$  of length  $p$  such that  $\pi' \in \Pi_{p-1}$  and  $a_{\pi'} = \min \mathbb{L}_{\mathcal{B}}(B_{\pi'}) < b_1$ . Then  $a_{\pi'} < b$  for every  $b \in \wr b_1, \dots, b_n \wr$ . For such a path  $\pi \in \Pi_p$ , we define

- $a_\pi := (a_{\pi'}, b_{\pi_p})$ ,
- $B_\pi := J(B_{\pi'}, a_{\pi'}, b_{\pi_p})$ .

Then  $a_\pi \in \mathcal{B}$  because  $\mathbb{L}_{\mathcal{B}}(B_{\pi'})$  is alphabetic and  $a_{\pi'} < b_{\pi_p}$ . The first, fourth and fifth claimed properties above are obviously satisfied. The second one is a consequence of Lemma 3.6 because  $a_{\pi'} \neq b_{\pi_p}$ . We also have the third one because the right-sibling of  $b_1$  in  $B_\pi$  is either the right sibling of  $b_1$  in  $B_{\pi'}$ , let us call it  $\tilde{b} \in \mathcal{B}$ , or  $(a_{\pi'}, \tilde{b})$  and in this case  $\tilde{b} = b_{\pi_p}$  thus  $(a_{\pi'}, \tilde{b}) = (a_{\pi'}, b_{\pi_p}) = a_\pi \in \mathcal{B}$ . See Appendix A for an example of the construction of  $B_\pi$ .

*Jacobi identity and size estimate.* We now prove that, for each  $p \in \llbracket 0, n-1 \rrbracket$  and  $\pi \in \Pi_p$ , there holds

$$\|\mathbf{I}(B_\pi)\| \leq (n-p-1)! + \sum_{\bar{\pi} \in S(\pi)} \|\mathbf{I}(B_{\bar{\pi}})\|, \quad (4.9)$$

where

$$S(\pi) := \{\bar{\pi} \in \Pi_{p+1}; \bar{\pi}' = \pi\}, \quad (4.10)$$

where, by convention,  $\Pi_n := \emptyset$ . Let  $p \in \llbracket 0, n-1 \rrbracket$  and  $\pi \in \Pi_p$ . Then either  $b_1$  or  $a_\pi$  is a minimal element of  $L_{\mathcal{B}}(B_\pi)$ .

- Assume that  $b_1$  is a minimal element of the leaves of  $B_\pi$ . Let  $\ell \in \mathcal{B}$  be its sibling leaf ( $\ell$  is either some  $b_k$  for  $k \neq 1$  or  $a_\pi$ ). Then  $(b_1, \ell) \in \mathcal{B}$ . Let  $t := J(B_\pi, b_1, \ell)$ . Then  $\mathbf{I}(t) = \mathbf{I}(B_\pi)$  and  $|t|_{\mathcal{B}} = |B_\pi|_{\mathcal{B}} - 1 = n - p$ . By Lemma 3.6,  $L_{\mathcal{B}}(t)$  is still alphabetic. By Theorem 3.10,  $\|\mathbf{I}(t)\| \leq (n-p-1)!$ .
- Otherwise,  $a_\pi$  is a minimal element of the leaves of  $B_\pi$ .
  - Assume that the sibling of  $a_\pi$  in  $B_\pi$  is a leaf  $b_k$ . Then  $(a_\pi, b_k) \in \mathcal{B}$ . Let  $t := J(B_\pi, a_\pi, b_k)$ . Then  $\mathbf{E}(\mathbf{I}(t)) = \pm \mathbf{E}(\mathbf{I}(B_\pi))$  in  $\mathcal{L}(X)$  and  $|t|_{\mathcal{B}} = |B_\pi|_{\mathcal{B}} - 1 = n - p$ . By Lemma 3.6,  $L_{\mathcal{B}}(t)$  is still alphabetic. By Theorem 3.10,  $\|\mathbf{I}(t)\| \leq (n-p-1)!$ .
  - Otherwise, the sibling of  $a_\pi$  is a tree  $w$ . Two cases occur.
    - \* First, if  $b_1$  is not a leaf of  $w$ , using the iterated Jacobi identity, we have in  $\mathcal{L}(X)$ ,  $\mathbf{E}(\mathbf{I}(B_\pi)) = \sum_{\bar{\pi} \in S(\pi)} \pm \mathbf{E}(\mathbf{I}(B_{\bar{\pi}}))$  so  $\|\mathbf{I}(B_\pi)\| \leq \sum_{\bar{\pi} \in S(\pi)} \|\mathbf{I}(B_{\bar{\pi}})\|$ .
    - \* Second, if  $b_1$  is a leaf of  $w$ , using the iterated Jacobi identity, we have in  $\mathcal{L}(X)$ ,  $\mathbf{E}(\mathbf{I}(B_\pi)) = \pm \mathbf{E}(\mathbf{I}(t_\pi)) + \sum_{\bar{\pi} \in S(\pi)} \pm \mathbf{E}(\mathbf{I}(B_{\bar{\pi}}))$  so  $\|\mathbf{I}(B_\pi)\| \leq \|\mathbf{I}(t_\pi)\| + \sum_{\bar{\pi} \in S(\pi)} \|\mathbf{I}(B_{\bar{\pi}})\|$ , where  $t_\pi := J(B_\pi, a_\pi, b_1)$  and  $|t_\pi|_{\mathcal{B}} = |B_\pi|_{\mathcal{B}} - 1 = n - p$  so  $\|\mathbf{I}(t_\pi)\| \leq (n-p-1)!$

*Final estimate.* Iterating (4.9) proves that

$$\|i(B_\emptyset)\| \leq \sum_{p=0}^{n-1} \sum_{\pi \in \Pi_p} (n-p-1)! \leq \sum_{p=0}^{n-1} \binom{n-1}{p} (n-p-1)! = \sum_{p=0}^{n-1} \frac{(n-1)!}{p!} = \lfloor e(n-1)! \rfloor \quad (4.11)$$

where the last equality holds for  $n \geq 2$  (see e.g. sequence A000522 in OEIS).  $\square$

In the next paragraph, we prove the optimality of estimate (4.7), by forcing both inequalities in (4.11) to be equalities: we construct a Hall set and candidates  $a, b$  for which any admissible path is in the last configuration (i.e.  $a_\pi$  is a minimal element of  $\mathbb{L}_B(B_\pi)$ , the sibling of  $a_\pi$  in  $B_\pi$  is a tree and  $b_1$  is a leaf of this tree), and the number of the admissible paths of length  $p$  is exactly  $\binom{n-1}{p}$ .

#### 4.4 Optimality case

Let  $n \geq 2$ . We prove that (4.7) is sharp with  $\theta_a(b) = n$ . Since we chose  $b$  such that  $|b| = n$ , our construction proves that the length-based estimate (1.3) of the introduction is also sharp. We start by constructing an order and the associated Hall set, then we exhibit brackets for which the estimate is saturated.

Let  $\overline{X} := \{X_2, \dots, X_n\}$  and  $X := \{X_0, X_1\} \cup \overline{X}$ . Let

$$\Pi := \bigcup_{0 \leq p \leq n-1} \{\pi = (\pi_1, \dots, \pi_p) \in \llbracket 2, n \rrbracket^p; \pi_1 < \dots < \pi_p\}, \quad (4.12)$$

the set of multi-indexes in increasing order. For  $\pi \in \Pi$ , we define

- $a_\pi := (\dots((X_0, X_{\pi_1}), X_{\pi_2}), \dots, X_{\pi_p})$ ,
- $y_\pi := (a_\pi, X_1)$ .

In particular,  $a_\emptyset = X_0$  and  $y_\emptyset = (X_0, X_1)$ . By Lemma 1.19,  $\text{Br}(\overline{X})$  is isomorphic to  $\text{Br}_{\overline{X}}$ , the submagma of  $\text{Br}(X)$  generated by  $\overline{X}$ . Let  $\text{Br}_\Pi$  denote the subset of  $\text{Br}(X)$  made of brackets involving exactly one  $y_\pi$  and any additional number of indeterminates of  $\overline{X}$ . ( $\text{Br}_\Pi$  is well-defined, since an equivalent characterization is that it is subset of trees  $t \in \text{Br}(X)$  involving  $X_0$  and  $X_1$  exactly once, within which the sibling of  $X_1$  is of the form  $a_\pi$  for some  $\pi \in \Pi$  and all other leaves are in  $\overline{X}$ ).

**Lemma 4.13.** *There exists a total Hall order  $<$  on  $\text{Br}(X)$  such that*

$$X_0 < X_1 < X_2 < \dots < X_n, \quad (4.13)$$

$$\forall j \in \llbracket 2, n-1 \rrbracket, \quad (\dots(X_1, X_n) \dots, X_{j+1}) < X_j \quad (4.14)$$

$$\forall \pi \in \Pi, \quad a_\pi < X_1, \quad (4.15)$$

$$\forall c_1, c_2 \in \text{Br}(\overline{X}), \quad \max L_{\overline{X}}(c_1) < \max L_{\overline{X}}(c_2) \Rightarrow c_1 < c_2, \quad (4.16)$$

$$\forall c_1 \in \text{Br}(\overline{X}), \forall c_2 \in \text{Br}_\Pi, \quad c_1 < c_2. \quad (4.17)$$

*Proof.* Our strategy consists in defining a Hall order on a  $\lambda$ -stable subset  $A$  of  $\text{Br}(X)$ , ensuring the desired conditions, and then extend it to  $\text{Br}(X)$  using Proposition 1.33. Let  $A := A_1 \cup A_2 \cup A_3 \cup A_4$ , where  $A_1, A_2, A_3, A_4$  are the pairwise disjoint sets

$$\begin{aligned} A_1 &:= \{a_\pi; \pi \in \Pi\}, & A_2 &:= \{X_1\} \cup \{(\dots(X_1, X_n) \dots, X_j); j \in \llbracket 2, n \rrbracket\}, \\ A_3 &:= \text{Br}(\overline{X}), & A_4 &:= \text{Br}_\Pi. \end{aligned} \quad (4.18)$$

Let  $\prec$  be any Hall order on  $\text{Br}(X)$ . We will use  $\prec$  as an arbitrary order to compare brackets for which we have no particular requirement.

*Step 1: We define a total order  $<$  on  $A$  in the following way.*

- (O1) If  $c_1 \in A_i$  and  $c_2 \in A_j$  with  $i \neq j$ ,  $c_1 < c_2$  if and only if  $i < j$  (i.e.  $A_1 < A_2 < A_3 < A_4$ )
- (O2) If  $c_1 \neq c_2 \in A_1$  (resp.  $A_2, A_4$ ) then  $c_1 < c_2$  if and only if  $c_1 \prec c_2$ .
- (O3) If  $c_1 \neq c_2 \in A_3 = \text{Br}(\overline{X})$ ,  $c_1 < c_2$  if and only if  $\max L_{\overline{X}}(c_1) < \max L_{\overline{X}}(c_2)$  or  $\max L_{\overline{X}}(c_1) = \max L_{\overline{X}}(c_2)$  and  $c_1 \prec c_2$ , where the maximum is computed for the natural ordering  $X_2 < \dots < X_n$  of  $\overline{X}$ .

One easily checks that the relation  $<$  is indeed transitive (due to its lexicographic nature) and a total order on  $A$  (thanks to the use of  $\prec$  as a last resort to break equalities).

*Step 2:* We check that  $A$  is  $\lambda$ -stable and that  $<$  is a Hall order on  $A$ , i.e. for any  $c \in A \setminus X$ ,  $\lambda(c) \in A$  and  $\lambda(c) < c$ .

- Let  $c \in A_1 \setminus X = A_1 \setminus \{X_0\}$ . Then  $c = a_\pi$  with  $\pi \neq \emptyset$  so  $\lambda(c) = a_{\pi'} \in A_1$ . Hence, since  $\prec$  is a Hall order, we have  $\lambda(c) \prec c$ , and by definition of  $<$  on  $A_1$ ,  $\lambda(c) < c$ .
- Let  $c \in A_2 \setminus X = A_2 \setminus \{X_1\}$ . Then  $\lambda(c) \in A_2$ , so  $\lambda(c) \prec c$  so  $\lambda(c) < c$ .
- Let  $c \in A_3 \setminus X = \text{Br}(\overline{X}) \setminus \overline{X}$ . Then  $\lambda(c) \in \text{Br}(\overline{X})$ . Moreover  $\max L_{\overline{X}}(\lambda(c)) \leq \max L_{\overline{X}}(c)$  because each leaf of  $\lambda(c)$  is also a leaf of  $c$ . So, by definition of  $<$  on  $A_3$ , either  $\max L_{\overline{X}}(\lambda(c)) < \max L_{\overline{X}}(c)$  and  $\lambda(c) < c$ , or  $\max L_{\overline{X}}(\lambda(c)) = \max L_{\overline{X}}(c)$  and  $\lambda(c) < c$  if and only if  $\lambda(c) \prec c$ , which is the case because  $\prec$  is a Hall order on  $\text{Br}(\overline{X})$ .
- Let  $c \in A_4 = \text{Br}_\Pi$ . First, if  $c = y_\pi$  for some  $\pi \in \Pi$ , then  $\lambda(y_\pi) = a_\pi \in A_1$  and  $\lambda(c) < c$  by (O1). Then,
  - either  $\lambda(c) \in \text{Br}(\overline{X}) = A_3$ , in which case  $\lambda(c) < c$  by (O1),
  - or  $\lambda(c) \in \text{Br}_\Pi$ , in which case  $\lambda(c) < c$  if and only if  $\lambda(c) \prec c$  (see (O2)), which is the case since  $\prec$  is a Hall order on  $\text{Br}(\overline{X})$ .

*Step 3: Conclusion.* By Proposition 1.33,  $<$  can be extended as a Hall order on  $\text{Br}(X)$ . Then (O1) ensures (4.13), (4.14), (4.15) and (4.17) while (O3) ensures (4.16), which concludes the proof. The purpose of (O2) is to fill in the unconstrained comparisons to build a total Hall order on  $A$ .  $\square$

**Proposition 4.14.** *Let  $<$  be the order constructed in Lemma 4.13 and  $\mathcal{B} \subset \text{Br}(X)$  the associated Hall set. Then  $b := (\dots((X_1, X_n), X_{n-1}), \dots, X_2) \in \mathcal{B}$  and  $\|[X_0, b]\|_{\mathcal{B}} = \lfloor e(n-1)! \rfloor$ .*

*Proof. Step 1:* We prove that  $b \in \mathcal{B}$ . By (4.13),  $(X_1, X_n) \in \mathcal{B}$ . Using (4.14), we prove by induction on  $j \in \{n-1, \dots, 2\}$  that  $(\dots, (X_1, X_n) \dots, X_j) \in \mathcal{B}$ . The case  $j = 2$  gives  $b \in \mathcal{B}$ .

*Step 2:* We prove that, for any  $\pi \in \Pi$ , then  $a_\pi, y_\pi \in \mathcal{B}$ . By definition of  $y_\pi$  and (4.15), it is sufficient to prove that  $a_\pi \in \mathcal{B}$ , which can be obtained by induction on the length of  $\pi$ , because  $a_\pi = (a_{\pi'}, X_{\pi_p})$  and  $a_{\pi'} < X_{\pi_p}$  by (4.15) and (4.13).

From now on, we use the same vocabulary as in the proof of Theorem 4.12, with  $a \leftarrow X_0$  and  $b_j \leftarrow X_j$  for  $j \in \llbracket 1, n \rrbracket$ .

*Step 3:* Any  $\pi \in \Pi$  is an admissible path. The proof is by induction on its length  $p \in \llbracket 0, n-1 \rrbracket$  and trivially holds for  $p = 0$ . When  $p \geq 1$  then  $\pi'$  is an admissible path by the induction assumption and  $\mathbb{L}_{\mathcal{B}}(B_{\pi'}) = \lfloor a_{\pi'}, X_1 \rfloor + \lfloor X_j; j \in \llbracket 2, n \rrbracket \setminus \pi \rfloor$  thus  $a_{\pi'} = \min \mathbb{L}_{\mathcal{B}}(B_{\pi'}) < X_1$  by (4.15).

For  $\pi \in \Pi$ , we define  $t_\pi, \tilde{t}_\pi \in \text{Br}(\mathcal{B})$  by

$$t_\pi = \langle \dots \langle y_\pi, X_{j_{n-p-1}} \rangle \dots, X_{j_1} \rangle, \quad (4.19)$$

$$\tilde{t}_\pi = \langle \dots \langle \langle X_1, a_\pi \rangle, X_{j_{n-p-1}} \rangle \dots, X_{j_1} \rangle, \quad (4.20)$$



where  $\{j_1 < \dots < j_{n-p-1}\} \cup \{\pi_1 < \dots < \pi_p\}$  is a partition of  $\llbracket 2, n \rrbracket$ . Then  $\mathbb{E}(\mathbb{I}(\tilde{t}_\pi)) = -\mathbb{E}(\mathbb{I}(t_\pi))$ .

*Step 4:* We prove the following equality in  $\mathcal{L}(X)$ .

$$[a, b] = \sum_{\pi \in \Pi, \pi_p \leq n-1} \pm \mathbb{E}(\mathbb{I}(t_\pi)) + \sum_{\pi \in \Pi, \pi_p = n} \pm \mathbb{E}(\mathbb{I}(\tilde{t}_\pi)) = \sum_{\pi \in \Pi} \pm \mathbb{E}(\mathbb{I}(t_\pi)). \quad (4.21)$$

Let  $\pi \in \Pi$ . If  $p = |\pi| \geq 1$  and  $\pi_p = n$ , then  $B_\pi$  is of the form  $\tilde{t}_\pi$  so  $\mathbb{I}(B_\pi) = \mathbb{I}(\tilde{t}_\pi)$ . Otherwise, the sibling of  $a_\pi$  in  $B_\pi$  is a non-trivial tree having  $X_1$  as a leaf. Therefore (see the last step in the proof of Theorem 4.12), we have in  $\mathcal{L}(X)$ ,  $\mathbb{E}(\mathbb{I}(B_\pi)) = \pm \mathbb{E}(\mathbb{I}(t_\pi)) + \sum_{\bar{\pi} \in \mathcal{S}(\pi)} \pm \mathbb{E}(\mathbb{I}(B_{\bar{\pi}}))$ . By iterating this relation starting from  $B_\emptyset$ , we obtain the conclusion (4.21).

*Step 5:* We prove that, for any  $\pi \in \Pi$ ,  $\|\mathbb{I}(t_\pi)\|_{\mathcal{B}} = (n-p-1)!$  and  $\text{supp } \mathbb{E}(\mathbb{I}(t_\pi)) \subset \mathbb{I}(\text{Tr}(\mathbb{L}_{\mathcal{B}}(t_\pi)))$ . Let  $\pi \in \Pi$ . We already know that the set  $\Delta := L_{\mathcal{B}}(t_\pi) = \{X_{j_1}, \dots, X_{j_{n-p-1}}, y_\pi\}$  is alphabetic. Moreover,  $X_{j_1} < \dots < X_{j_{n-p-1}} < y_\pi$  by (4.13) and (4.17). By Lemma 1.42,  $\Delta$  is free because  $\Delta \subset \mathcal{B} \cap \text{Br}_X$  and  $a_\pi = \lambda(y_\pi) < \min(\Delta)$  (see (4.13) and (4.15)).

Let us check that, for every  $c_1, c_2 \in \mathcal{B} \cap \text{Br}_\Delta$  with  $L_\Delta(c_1) \cap L_\Delta(c_2) = \emptyset$ , then  $c_1 < c_2$  iff  $\max L_\Delta(c_1) < \max L_\Delta(c_2)$ . If  $c_1, c_2 \in \text{Br}(\{X_{j_1}, \dots, X_{j_{n-p-1}}\})$  then (4.16) gives the conclusion. Otherwise,  $y_\pi$  is a leaf of (exactly) one of the two elements  $c_1, c_2$  so (4.17) gives the conclusion.

Hence, the conclusion of this step follows from Proposition 3.12 (and Theorem 3.10 for the structural part).

*Step 6:* We prove that the sets  $\mathcal{B} \cap \mathbb{I}(\text{Tr}(\mathbb{L}_{\mathcal{B}}(t_\pi)))$ , for  $\pi \in \Pi$ , are two by two disjoint. Let  $\pi \neq \pi^b \in \Pi$ ,  $c \in \mathcal{B} \cap \mathbb{I}(\text{Tr}(\mathbb{L}_{\mathcal{B}}(t_\pi)))$  and  $c^b \in \mathcal{B} \cap \mathbb{I}(\text{Tr}(\mathbb{L}_{\mathcal{B}}(t_{\pi^b})))$ . Then  $c = \mathbb{I}(B)$  and  $c^b = \mathbb{I}(B^b)$ , where  $B$  and  $B^b$  are two different elements of  $\text{Br}(\Delta)$  where  $\Delta := \{y_\pi, y_{\pi^b}\} \cup \{X_j; j \in \llbracket 2, n \rrbracket\}$ . By Lemma 1.42 (with  $A = X$ ),  $\Delta$  is free because  $\Delta \subset \mathcal{B} \cap \text{Br}_X$ ,  $a_\pi = \lambda(y_\pi) < \min(\Delta)$  and  $a_{\pi^b} = \lambda(y_{\pi^b}) < \min(\Delta)$  (see (4.13) and (4.15)). Thus  $\mathbb{I} : \text{Br}(\Delta) \rightarrow \text{Br}_\Delta$  is injective and  $c \neq c^b$ .

We deduce from Steps 4, 5 and 6 that

$$\|[X_0, b]\|_{\mathcal{B}} = \sum_{\pi \in \Pi} \|\mathbb{I}(t_\pi)\|_{\mathcal{B}} = \sum_{p=0}^{n-1} \binom{n-1}{p} (n-p-1)! = \lfloor e(n-1)! \rfloor, \quad (4.22)$$

which concludes the proof.  $\square$

The example constructed above is stated in a finite setting. However, straightforward adaptations lead to the following consequence for infinite sets of indeterminates.

**Corollary 4.15.** *Assume that  $X$  is infinite. There exists a Hall set  $\mathcal{B} \subset \text{Br}(X)$  such that, for every  $n \geq 2$ , there exists  $a < b \in \mathcal{B}$  with  $|a| = 1$ ,  $\theta_a(b) = |b| = n$  and  $\|[a, b]\|_{\mathcal{B}} = \lfloor e(n-1)! \rfloor$ .*

*Proof.* We can assume that  $X$  contains indeterminates labeled as  $X_i$  for  $i \in \mathbb{N}$ . We then perform the following slight modifications to the construction detailed above. First,  $\Pi$  defined in (4.12) is changed to  $\Pi := \bigcup_{p \in \mathbb{N}} \{\pi = (\pi_1, \dots, \pi_p); 2 \leq \pi_1 < \dots < \pi_p\}$ . Then, we prove that there exists a total Hall order on  $\text{Br}(X)$  satisfying the all conditions of Lemma 4.13 (where (4.13) and (4.14) are to be understood as holding for every  $n \geq 2$ ). This solely requires to modify the set  $A_2$  of (4.18) into  $A_2 := \{X_1\} \cup \{(\dots (X_1, X_n) \dots, X_j); j \in \llbracket 2, n \rrbracket; n \geq 2\}$ . Then the result follows by considering the Hall set associated with this order and the brackets of Proposition 4.14.  $\square$

## 4.5 Recursion depth of the decomposition algorithm

In this paragraph, we prove that  $\theta_a(b)$  is strictly decreasing along the historical recursive decomposition algorithm described in Section 2.1, within any Hall set  $\mathcal{B} \subset \text{Br}(X)$ .

Hence, from a computational point of view, although this algorithm is classically seen as an induction on the couple  $(|a| + |b|, \min\{a, b\})$ , this shows that the size of the associated call stack

(number of simultaneous nested executions) is bounded by  $\theta_a(b) \leq |b|$ . This is not straightforward, since it is not guaranteed that  $|b|$  or  $|a| + |b|$  decrease along the iterations.

From a theoretical point of view, we will use this property to prove results relying on the recursive decomposition algorithm by induction on  $\theta_a(b)$  for some particular Hall sets.

We start with a technical lemma which states a kind of stability of  $\theta_a(b)$  with respect to  $b$  within a certain class (monotony with respect to  $a$  is covered in Lemma 4.8).

**Lemma 4.16.** *Let  $a < b \in \mathcal{B}$  and  $c \in \text{supp}[a, b]$ . Then  $\theta_a(c) = \theta_a(b)$  and  $\lambda(c) \neq b$ .*

*Proof.* Let  $a < b \in \mathcal{B}$ . We start by proving two elementary claims, which lead to the result.

*Step 1:* For each  $t \in \text{Br}(L_{\mathcal{B}}(T_a(b)))$  with  $\mathfrak{I}(t) \in \mathcal{B}$ , one has  $\theta_a(\mathfrak{I}(t)) = |t|_{\mathcal{B}}$ .

We proceed by induction on  $|t|_{\mathcal{B}}$ . For each such  $t$ , by Lemma 1.41,  $a < \mathfrak{I}(t)$ . When  $|t|_{\mathcal{B}} = 1$ ,  $t$  is of the form  $b_i$  for some  $b_i \in L_{\mathcal{B}}(T_a(b))$  so  $(a, \mathfrak{I}(t)) = (a, b_i) \in \mathcal{B}$  hence  $\theta_a(\mathfrak{I}(t)) = 1$ . When  $|t|_{\mathcal{B}} > 1$ ,  $\lambda(t) \in \text{Br}(L_{\mathcal{B}}(T_a(b)))$  so  $a < \lambda(\mathfrak{I}(t))$ . Hence  $\theta_a(\mathfrak{I}(t)) = \theta_a(\mathfrak{I}(\lambda(t))) + \theta_a(\mathfrak{I}(\mu(t)))$  and the conclusion follows by induction.

*Step 2:* For each  $t \in \text{Br}(\{a\} \cup L_{\mathcal{B}}(T_a(b)))$  with  $\mathfrak{I}(t) \in \mathcal{B}$ ,  $|t|_{\mathcal{B}} \geq 2$  involving  $a$  exactly once,  $\theta_a(\mathfrak{I}(t)) = |t|_{\mathcal{B}} - 1$ .

We also proceed by induction on  $|t|_{\mathcal{B}}$ . When  $|t|_{\mathcal{B}} = 2$ ,  $t$  is of the form  $\langle a, b_i \rangle$  for some  $b_i \in L_{\mathcal{B}}(T_a(b))$ , so  $\mathfrak{I}(t) = (a, b_i)$  and  $\lambda(\mathfrak{I}(t)) = a$ , hence  $\theta_a(\mathfrak{I}(t)) = 1$ . When  $|t|_{\mathcal{B}} > 2$ , one cannot have  $\lambda(t) = a$ . Indeed, if  $\lambda(t) = a$ , since  $|\mu(t)|_{\mathcal{B}} \geq 2$ ,  $\lambda(\mu(t)) \in \text{Br}(L_{\mathcal{B}}(T_a(b)))$  so  $a < \lambda(\lambda(\mu(t)))$ , which contradicts the fact that  $\mathfrak{I}(t) \in \mathcal{B}$ . Hence,  $\theta_a(\mathfrak{I}(t)) = \theta_a(\mathfrak{I}(\lambda(t))) + \theta_a(\mathfrak{I}(\mu(t)))$ , where  $\lambda(t)$  and  $\mu(t)$  satisfy respectively the hypotheses of Step 1 and Step 2 (or the converse). Thus the conclusion follows by induction.

*Step 3: Conclusion.* Let  $c \in \text{supp}[a, b]$ . By Theorem 4.12,  $c \in \mathfrak{I}(\text{Tr}(\mathbb{A}))$  where  $\mathbb{A} = \{a\} \cup L_{\mathcal{B}}(T_a(b))$ . Hence, by the previous argument,  $\theta_a(c) = (\theta_a(b) + 1) - 1 = \theta_a(b)$ . By contradiction, if  $\lambda(c) = b$ , then one would have  $\theta_a(c) = \theta_a(\lambda(c)) + \theta_a(\mu(c)) > \theta_a(b)$ .  $\square$

**Proposition 4.17.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. Let  $\text{Rewrite}(\cdot, \cdot)$  denote the algorithm described in Section 2.1 which associates to each couple  $a < b \in \mathcal{B}$  the decomposition of  $[a, b]$  on  $\mathcal{B}$ . For every  $a < b \in \mathcal{B}$ , calling  $\text{Rewrite}(a, b)$  requires a call stack of size at most  $\theta_a(b)$ .*

*Proof.* We proceed by induction on  $n := \theta_a(b)$ . First, when  $n = \theta_a(b) = 1$ ,  $(a, b) \in \mathcal{B}$  so there is no nested call. Let  $n \geq 2$  and  $a < b \in \mathcal{B}$  such that  $\theta_a(b) = n$ . We need to check that  $\text{Rewrite}(a, b)$  only calls  $\text{Rewrite}(\tilde{a}, \tilde{b})$  with arguments  $\tilde{a} < \tilde{b}$  such that  $\theta_{\tilde{a}}(\tilde{b}) < \theta_a(b)$ . Since  $\theta_a(b) > 1$ ,  $b \notin X$  and  $a < \lambda(b) < \mu(b)$ . Jacobi's identity yields

$$[a, b] = [[a, \lambda(b)], \mu(b)] + [\lambda(b), [a, \mu(b)]]. \quad (4.23)$$

We study both terms separately, but with the same method.

- *First term.* Calling  $\text{Rewrite}(a, \lambda(b))$ , we obtain  $[a, \lambda(b)] = \sum \alpha_d \mathfrak{E}(d)$ . Moreover, one has  $\theta_a(\lambda(b)) < \theta_a(b)$  (by definition, because  $\theta_a(\mu(b)) \geq 1$ ). For each  $d$  in this decomposition,
  - If  $d < \mu(b)$ , we call  $\text{Rewrite}(d, \mu(b))$ . By (2.2),  $a \leq \lambda(d) < d$ . Hence, by Lemma 4.8,  $\theta_d(\mu(b)) \leq \theta_a(\mu(b)) < \theta_a(b)$ .
  - If  $\mu(b) < d$ , we call  $\text{Rewrite}(\mu(b), d)$ . By Lemma 4.8,  $\theta_{\mu(b)}(d) \leq \theta_a(d)$ . Moreover, since  $d \in \text{supp}[a, \lambda(b)]$ , by Lemma 4.16,  $\theta_a(d) \leq \theta_a(\lambda(b))$ . Hence  $\theta_{\mu(b)}(d) < \theta_a(b)$ .
- *Second term.* Calling  $\text{Rewrite}(a, \mu(b))$ , we obtain  $[a, \mu(b)] = \sum \alpha_d \mathfrak{E}(d)$ . Moreover, one has  $\theta_a(\mu(b)) < \theta_a(b)$  (by definition, because  $\theta_a(\lambda(b)) \geq 1$ ). For each  $d$  in this decomposition,
  - If  $d < \lambda(b)$ , we call  $\text{Rewrite}(d, \lambda(b))$ . By (2.2),  $a \leq \lambda(d) < d$ . Hence, by Lemma 4.8,  $\theta_d(\lambda(b)) \leq \theta_a(\lambda(b)) < \theta_a(b)$ .

- If  $\lambda(b) < d$ , we call  $\text{Rewrite}(\lambda(b), d)$ . By Lemma 4.8,  $\theta_{\lambda(b)}(d) \leq \theta_a(d)$ . Moreover, since  $d \in \text{supp}[a, \mu(b)]$ , by Lemma 4.16,  $\theta_a(d) \leq \theta_a(\mu(b))$ . Hence  $\theta_{\lambda(b)}(d) < \theta_a(b)$ .

This concludes the proof.  $\square$

**Remark 4.18.** *With this approach, one could also estimate the size of the structure constants. By induction on  $n := \theta_a(b) \geq 1$ , one constructs a sequence  $C(n)$  such that  $\|[a, b]\|_{\mathcal{B}} \leq C(\theta_a(b))$ . If the worst cases happen simultaneously, the previous proof shows that one can take  $C(n) := 2C(n-1)^2$ . Hence  $C(n) := 2^{2^{n-1}-1}$  is an admissible bound. This yields an estimate of the form*

$$\|[a, b]\|_{\mathcal{B}} \leq 2^{2^{\theta_a(b)-1}-1}, \quad (4.24)$$

which, since  $\theta_a(b) \leq |b|$ , is of course much better than (2.3), but also very far from the previous optimal estimate (4.7), which stems from a tighter tracking of the unfavorable cases.

## 5 Systematic geometric lower bounds

The main goal of this section is to prove Theorem 1.4 which yields a geometric lower bound for the structure constants relative to any Hall set. In Section 5.1, we gather elementary results concerning systematic basis elements used in the next sections. We then prove Theorem 1.4 in Section 5.2 when  $|X| \geq 3$  (which is straightforward and easily checked to be optimal) and in Section 5.3 when  $|X| = 2$  (which relies on more involved computations). The optimality of the lower bound in this latter case is discussed in Section 8. Eventually, we prove in Section 5.4 a systematic  $\theta$ -based geometric lower-bound.

To measure the worst-case growth of the structure constants of  $\mathcal{L}(X)$  relative to a given Hall set  $\mathcal{B} \subset \text{Br}(X)$ , we will use the following quantity for  $n \geq 2$ :

$$\beta_n(\mathcal{B}) := \sup \{ \|[a, b]\|_{\mathcal{B}}; a < b \in \mathcal{B}, |a| + |b| = n \}. \quad (5.1)$$

### 5.1 Some linearly independent elements of any Hall set

We identify some elements which belong to a Hall set without knowing the underlying order.

**Proposition 5.1.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set.*

- *If  $\{X_0, X_1\} \subset X$ ,  $b_0 \in \mathcal{B}$  with  $b_0 < X_1$  and  $r := r(b_0, X_1)$  is defined by Lemma 4.6, then the following elements of  $\text{Br}(X)$  belong to  $\mathcal{B}$ :*
  1.  $b_k := \text{ad}_{X_1}^k(b_0)$  for  $1 \leq k \leq r$ ,
  2.  $\text{ad}_{X_1}^n(b_r)$  for  $n \in \mathbb{N}$ ,
  3.  $w_i := (b_i, b_{i+1}) = \text{ad}_{b_i}^2(X_1)$  for  $0 \leq i \leq r-1$ ,
  4.  $(\text{ad}_{X_1}^n(b_r), \text{ad}_{X_1}^m(b_r))$  for  $m, n \in \mathbb{N}$  such that  $\text{ad}_{X_1}^n(b_r) < \text{ad}_{X_1}^m(b_r)$ ,
  5. for  $0 \leq i \leq r-1$ , if  $w_i < X_1$ ,  $\text{ad}_{X_1}^n(w_i)$  for  $n \leq q_i := r(w_i, X_1)$  and  $\text{ad}_{X_1}^{n-q_i} \text{ad}_{X_1}^{q_i}(w_i)$  for  $n > q_i$ , otherwise  $\text{ad}_{X_1}^n(w_i)$  for all  $n \geq 0$ .
- *If  $\{X_0, X_1, X_2\} \subset X$  with  $X_0 < X_1 < X_2$ , denoting by  $b_i = \text{ad}_{X_1}^k(X_0)$  and  $r := r(X_0, X_1)$ , the following elements of  $\text{Br}(X)$  lie in  $\mathcal{B}$ :*
  1.  $\text{ad}_{X_1}^n(X_2)$  for  $n \in \mathbb{N}$ ,
  2.  $\text{ad}_{X_1}^n(b_i, X_2)$  for  $0 \leq i \leq r-1$ ,  $n \in \mathbb{N}$ ,
  3. either  $(\text{ad}_{X_1}^n(b_r), \text{ad}_{X_1}^m(X_2))$  or  $(\text{ad}_{X_1}^m(X_2), \text{ad}_{X_1}^n(b_r))$  when  $m, n \in \mathbb{N}$ .

*Proof.* Let us prove both parts.

- Items (1) and (2) are Lemma 4.6. Item (3) holds because, for  $0 \leq i \leq r-1$ ,  $b_i < X_1$ .  
For item (4), let  $m, n \in \mathbb{N}$  such that  $\text{ad}_{X_1}^m(b_r) < \text{ad}_{X_1}^n(b_r)$ . If  $n \geq 1$ , then  $\lambda(\text{ad}_{X_1}^n(b_r)) = X_1 \leq \text{ad}_{X_1}^m(b_r)$ , either by Hall order property when  $m \geq 1$ , or because  $b_r > X_1$  when  $m = 0$ , thus  $(\text{ad}_{X_1}^n(b_r), \text{ad}_{X_1}^m(b_r)) \in \mathcal{B}$ . If  $n = 0$ , then  $m \geq 1$  thus  $\lambda(\text{ad}_{X_1}^m(b_r)) = \lambda(b_r) = b_{r-1} \leq X_1 \leq \text{ad}_{X_1}^n(b_r)$  which gives the conclusion.  
For item (5), if  $w_i < X_1$ , it suffices to apply Lemma 4.6. Of  $X_1 < w_i$ , since  $\lambda(w_i) = b_i < X_1$  because  $i < r$ , so  $(X_1, w_i) \in \mathcal{B}$  and the same holds for  $\text{ad}_{X_1}^n(w_i)$  for all  $n \geq 0$ .
- Since  $X_1 < X_2$ , item (1) is clear. Concerning item (2), for  $0 \leq i \leq r-1$ , we have  $b_i < X_1 < X_2$  thus  $(b_i, X_2) \in \mathcal{B}$  and  $(X_1, (b_i, X_2)) \in \mathcal{B}$ , which gives the conclusion. For item (3), let  $m, n \in \mathbb{N}$ . We assume  $\text{ad}_{X_1}^n(b_r) < \text{ad}_{X_1}^m(X_2)$ . If  $m = 0$  then  $(\text{ad}_{X_1}^n(b_r), \text{ad}_{X_1}^m(X_2)) = (\text{ad}_{X_1}^n(b_r), X_2) \in \mathcal{B}$ . If  $m \geq 1$ , then  $\lambda(\text{ad}_{X_1}^m(X_2)) = X_1 \leq \text{ad}_{X_1}^n(b_r)$ , by Hall order property when  $n \geq 1$ , because  $b_r > X_1$  when  $n = 0$ , thus  $(\text{ad}_{X_1}^n(b_r), \text{ad}_{X_1}^m(X_2)) \in \mathcal{B}$ . The same reasoning gives the conclusion when  $\text{ad}_{X_1}^m(X_2) < \text{ad}_{X_1}^n(b_r)$ .  $\square$

## 5.2 The general case

We prove an optimal systematic lower bound when  $|X| \geq 3$ .

**Theorem 5.2.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. Assume that  $|X| \geq 3$  and let  $X_0 < X_1 < X_2 \in X$ . For every  $n \in \mathbb{N}$ ,  $\|[X_0, \text{ad}_{X_1}^n(X_2)]\|_{\mathcal{B}} \geq 2^n$  with equality if  $r(X_0, X_1) \in \{1, +\infty\}$ . Thus  $\beta_{n+2}(\mathcal{B}) \geq 2^n$ .*

*Proof.* Let  $r := r(X_0, X_1)$  be defined in Lemma 4.6. By Lemma B.2 (applied to the derivation  $\text{ad}_{X_1}$  on  $\mathcal{L}(X)$ ,  $\nu \leftarrow n$ ,  $b_1 \leftarrow X_0$  and  $b_2 \leftarrow X_2$ ),

$$[X_0, \text{ad}_{X_1}^n(X_2)] = \sum_{k=0}^n (-1)^k \binom{n}{k} \text{ad}_{X_1}^{n-k} [\text{ad}_{X_1}^k(X_0), X_2]. \quad (5.2)$$

Case  $n < r$ . Then (5.2) implies that

$$[X_0, \text{ad}_{X_1}^n(X_2)] = \sum_{k=0}^n \binom{n}{k} \text{ad}_{X_1}^{n-k} [\underline{\text{ad}}_{X_1}^k(X_0), X_2]. \quad (5.3)$$

By item (2) of the second point of Proposition 5.1, the elements of the right-hand side are (evaluations of) distinct elements of  $\mathcal{B}$ . Thus,  $\|[X_0, \text{ad}_{X_1}^n(X_2)]\|_{\mathcal{B}} = \sum_{k=0}^n \binom{n}{k} = 2^n$ . In particular, when  $r = +\infty$  this equality holds for every  $n \in \mathbb{N}$ .

Case  $r$  finite and  $n \geq r$ . Starting from (5.2), using successively Jacobi's formula, the index change  $s = k + j$  and a re-arrangement, we obtain

$$\begin{aligned} [X_0, \text{ad}_{X_1}^n(X_2)] &= \sum_{k=0}^{r-1} \binom{n}{k} \text{ad}_{X_1}^{n-k} [\underline{\text{ad}}_{X_1}^k(X_0), X_2] \\ &= \sum_{k=r}^n (-1)^k \binom{n}{k} \sum_{j=0}^{n-k} \binom{n-k}{j} [\text{ad}_{X_1}^{k+j}(X_0), \text{ad}_{X_1}^{n-k-j}(X_2)] \\ &= \sum_{s=r}^n \alpha_s [\text{ad}_{X_1}^{s-r} \underline{\text{ad}}_{X_1}^r(X_0), \text{ad}_{X_1}^{n-s}(X_2)], \end{aligned} \quad (5.4)$$

where, using [18, (1.5)] for the last equality,

$$\alpha_s := (-1)^r \sum_{k=r}^s (-1)^k \binom{n}{k} \binom{n-k}{s-k} = (-1)^r \binom{n}{s} \sum_{k=r}^s (-1)^k \binom{s}{k} = \binom{n}{s} \binom{s-1}{r-1}. \quad (5.5)$$

By item (3) of the second point of Proposition 5.1, the elements in the right-hand side of (5.4) are, up to a sign which depends on  $\mathcal{B}$ , (evaluations of) distinct elements of  $\mathcal{B}$ . Thus  $\|[X_0, \text{ad}_{X_1}^n(X_2)]\|_{\mathcal{B}} = \sum_{k=0}^{r-1} \binom{n}{k} + \sum_{s=r}^n \alpha_s \geq 2^n$ , because  $\alpha_s \geq \binom{n}{s}$ , with equalities when  $r = 1$ .  $\square$

**Remark 5.3.** *This lower bound is optimal, in the sense that there exist Hall sets  $\mathcal{B} \subset \text{Br}(X)$  such that  $\beta_{n+2}(\mathcal{B}) \leq 2^n$ . In particular, this is the case of the length-compatible and Lyndon bases for which Theorem 1.5 and Theorem 1.6 prove that  $\|[a, b]\|_{\mathcal{B}} \leq 2^{|b|-1} \leq 2^{|a|+|b|-2}$ . Hence, for these bases  $\beta_{n+2}(\mathcal{B}) = 2^n$  when  $|X| \geq 3$ .*

### 5.3 The two-indeterminates case

We prove a systematic lower bound when  $|X| = 2$  in Theorem 5.5. The construction requires more complex brackets, for which we start by giving elementary formulas.

**Proposition 5.4.** *Let  $a, b \in \mathcal{L}(X)$ . For  $n \in \mathbb{N}$ , let  $b_n := \text{ad}_a^n(b)$ . For  $i, j \in \mathbb{N}$  with  $i \leq j$ , let  $w_{i,j} := [b_i, b_j]$  and  $w_i := w_{i,i+1} = [b_i, b_{i+1}]$ . The following relations hold for all  $i \in \mathbb{N}$  and  $n \in \mathbb{N}^*$*

$$w_{i,i+n} = \sum_{p=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^p \binom{n-1-p}{p} \text{ad}_a^{n-1-2p}(w_{i+p}), \quad (5.6)$$

$$\text{ad}_a^n(w_i) = \sum_{p=0}^{\lfloor \frac{n-1}{2} \rfloor} \left[ \binom{n}{p} - \binom{n}{p-1} \right] w_{i+p, i+1+n-p}. \quad (5.7)$$

The following relation holds for all  $n \in \mathbb{N}$  and  $r \in \llbracket 1, \lfloor \frac{n-1}{2} \rfloor \rrbracket$ ,

$$w_{0,n} = \sum_{p=0}^{r-1} (-1)^p \binom{n-1-p}{p} \text{ad}_a^{n-1-2p}(w_p) + \sum_{s=r}^{\lfloor \frac{n-1}{2} \rfloor} \left[ \sum_{p=r}^s (-1)^p \binom{n-1-p}{p} \left( \binom{n-1-2p}{s-p} - \binom{n-1-2p}{s-p-1} \right) \right] w_{s, n-s}. \quad (5.8)$$

*Proof.* The validity for every  $i \in \mathbb{N}$  of (5.6) and (5.7) is proved by induction on  $n \in \mathbb{N}^*$ . For  $n = 1$ , they hold because  $w_{i,i+1} = w_i$ . Let us prove the result for  $(n+1)$  knowing it for  $n$ .

For (5.6), we use the Jacobi relation  $w_{i,i+n+1} = [b_i, \text{ad}_a(b_{i+n})] = \text{ad}_a(w_{i,i+n}) - w_{i+1,i+n}$ , inject the induction assumption, re-organize terms and conclude with  $\binom{n-1-p}{p} + \binom{n-1-p}{p-1} = \binom{n-p}{p}$ .

For (5.7), we start from the induction assumption, use the Jacobi relation

$$\text{ad}_a(w_{i+p, i+1+n-p}) = w_{i+p+1, i+1+n-p} + w_{i+p, i+2+n-p}, \quad (5.9)$$

reorganize terms and conclude with  $\binom{n}{p} - \binom{n}{p-1} + \binom{n}{p-1} - \binom{n}{p-2} = \binom{n+1}{p} - \binom{n+1}{p-1}$ .

Finally, to prove (5.8), we start from (5.6) with  $i = 0$ , then split the sum in two sums over  $p \in \llbracket 0, r-1 \rrbracket$  and  $p \in \llbracket r, \lfloor \frac{n-1}{2} \rfloor \rrbracket$ . In each term of the second sum, we incorporate (5.7) and re-organize terms.  $\square$

We now prove the main theorem of this paragraph. We include here an upper bound, see (5.12) below, which is used in the proof of Proposition 6.11. We prove this upper bound here because it concerns the same bracket as the one which we use to prove systematic lower bounds.

**Theorem 5.5.** *Let  $X = \{X_0, X_1\}$ ,  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set such that  $X_0 < X_1$ ,  $r := r(X_0, X_1)$  defined by Lemma 4.6,  $b_n := \underline{\text{ad}}_{X_1}^n(X_0)$  for  $n \in \llbracket 0, r \rrbracket$  and  $b_n := \text{ad}_{X_1}^{n-r}(\underline{\text{ad}}_{X_1}^r(X_0))$  for  $n > r$ . Then,*

for every  $n \in \mathbb{N}^*$ ,

$$\| [X_0, b_n] \|_{\mathcal{B}} = F_n \quad \text{if } n < 2r + 1, \quad (5.10)$$

$$\| [X_0, b_n] \|_{\mathcal{B}} \geq \max\{F_n, C(r)n^{r-\frac{5}{2}}2^n\} \quad \text{if } n \geq 2r + 1, \quad (5.11)$$

$$\| [X_0, b_n] \|_{\mathcal{B}} \leq 2^{n-2} \quad \text{if } r = 1 \text{ and } n \geq 2. \quad (5.12)$$

where  $(F_\nu)_{\nu \in \mathbb{N}}$  denote the 0-based Fibonacci numbers. In particular  $\beta_{n+2}(\mathcal{B}) \geq F_n$ .

*Proof.* We use the notations of Proposition 5.4 with  $b_0 = X_0$ . If  $n < 2r + 1$ , then by item (5) of the first point of Proposition 5.1, the right-hand side elements of (5.6) with  $i = 0$  are, up to sign, evaluations of distinct elements of  $\mathcal{B}$ , so we have

$$\| [b_0, b_n] \|_{\mathcal{B}} = \sum_{p=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-1-p}{p} = F_n, \quad (5.13)$$

which proves (5.10). Now, we assume that  $r$  is finite and  $n \geq 2r + 1$ . By items (4) and (5) of the first point of Proposition 5.1, the right-hand side elements of (5.8) are, up to sign, evaluations of distinct elements of  $\mathcal{B}$ , so we have

$$\| [b_0, b_n] \|_{\mathcal{B}} = \sum_{p=0}^{r-1} \binom{n-1-p}{p} + \sum_{s=r}^{\lfloor \frac{n-1}{2} \rfloor} A_s^r(n) \quad (5.14)$$

where

$$A_s^r(n) := \left| \sum_{p=r}^s (-1)^p \binom{n-1-p}{p} \left( \binom{n-1-2p}{s-p} - \binom{n-1-2p}{s-p-1} \right) \right|. \quad (5.15)$$

Then (5.14) and Lemma C.3 prove that

$$\| [b_0, b_n] \|_{\mathcal{B}} \geq \sum_{p=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-1-p}{p} = F_n. \quad (5.16)$$

Moreover (5.14) and Lemma C.4 prove that

$$\| [b_0, b_n] \|_{\mathcal{B}} \geq A_{\lfloor \frac{n-1}{2} \rfloor}^r(n) \geq C_r n^{r-\frac{5}{2}} 2^n. \quad (5.17)$$

which concludes the proof of (5.11).

Eventually, the last estimate (5.12) follows from (5.14) with  $r = 1$  and Lemma C.5.  $\square$

**Remark 5.6.** *This estimate is optimal, in the sense that there exists a Hall set  $\mathcal{B} \subset \text{Br}(X)$  such that  $\beta_{n+2}(\mathcal{B}) = F_n$  for all  $n \geq 0$ . We construct such a Hall set in Section 8. Up to our knowledge, this Hall set has not been studied before. When  $|X| = 2$ , neither the length-compatible nor the Lyndon basis satisfy  $\beta_{n+2}(\mathcal{B}) = F_n$  (see Section 8.1 for more insight on this topic, and Corollary 6.12 and Proposition 7.8 for the computation of the exact values in the case of length-compatible and Lyndon bases).*

## 5.4 A $\theta$ -based lower bound

The previous paragraphs prove that the length-based lower bound is geometric in any Hall set. We prove that this also holds for the  $\theta$ -based lower bound, with a geometric ratio of 2, even in the two-indeterminates case.

**Proposition 5.7.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set with  $|X| \geq 2$ . For every  $\theta \in \mathbb{N}^*$  there exist  $a < b \in \mathcal{B}$  such that  $\theta_a(b) = \theta$  and  $\|[a, b]\|_{\mathcal{B}} \geq 2^{\theta-1}$ .*

*Proof.* Let  $X_0, X_1 \in X$  with  $X_0 < X_1$ . Now let  $\mathcal{X} := \{x_0, x_1, x_2\} = \{(X_0, X_1), (X_0, (X_0, X_1)), X_1\}$ , numbered such that  $x_0 < x_1 < x_2$ . The set  $\mathcal{X}$  is a free alphabetic subset of  $\mathcal{B}$ . Let  $\theta \in \mathbb{N}^*$ , and let  $b_\theta := \text{ad}_{x_1}^{\theta-1}(x_2)$ . Since  $x_0 < x_1 < x_2$ ,  $x_0 < b_\theta \in \mathcal{B}$ , and a straightforward induction on  $\theta$  shows that  $\theta_{x_0}(b_\theta) = \theta$ . By Proposition 3.5, the canonical map  $\mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(X)$  is an isometry with respect to the norms relative to  $\mathcal{B} \cap \text{Br}_{\mathcal{X}}$  and  $\mathcal{B}$  respectively. The fact that  $\|[x_0, b_\theta]\|_{\mathcal{B}} \geq 2^{\theta-1}$  is then a direct consequence of Theorem 5.2 applied in  $\mathcal{L}(\mathcal{X})$ .  $\square$

**Remark 5.8.** *As we will see in forthcoming sections, this lower bound of  $2^{\theta-1}$  for the  $\theta$ -based estimate is attained in the case of the length-compatible Hall sets (see Theorem 6.7) and in the case of the Lyndon basis (see Theorem 7.5). However, none of these Hall sets attains the lower bound for the length-based estimate in the case  $|X| = 2$ .*

*When  $|X| = 2$ , we will construct in Section 8 another Hall set which does not attain the lower bound for the  $\theta$ -based estimate, but attains the lower bound for the length-based estimate.*

## 6 Length-compatible Hall sets

The main goal of this section is to prove Theorem 1.5, which yields a sharp bound of structure constants relative to length-compatible Hall sets.

**Definition 6.1** (Length-compatible Hall sets). *We say that  $\mathcal{B}$  is a length-compatible Hall set when its order is length-compatible, i.e. satisfies*

$$\forall a, b \in \mathcal{B}, \quad a < b \Rightarrow |a| \leq |b|. \quad (6.1)$$

Since (6.1) does not impose the order between elements of the same length, which then determines whether some brackets belong or not the Hall set, there are many different length-compatible Hall sets, even for a fixed  $X$ . The lower and upper bounds we prove in this section are valid for all length-compatible Hall sets without distinction, highlighting that, with respect to our focus on worst-case growth of structure constants, they behave similarly.

We start with a short terminology discussion in Section 6.1 and preliminary results concerning structure constants for such bases in Section 6.2. We then prove an enhanced version of Theorem 1.5 in Section 6.3. Eventually, we discuss the optimality of length-focused estimates in Section 6.4.

### 6.1 Naming length-compatible Hall sets

Historically, Hall sets were introduced by Marshall Hall in [20], from ideas implicit in the work [21] of Philip Hall concerning group theory. In Marshall Hall's definition, the third condition in Definition 1.26 was replaced by the stronger condition (6.1). Since one always has  $|a| < |a| + |b|$ , condition (6.1) implies that  $a < (a, b)$  so it is indeed stronger.

**Remark 6.2.** *Some authors refer to such length-compatible sets as “Philip Hall bases” (e.g. [42]) or “Philip Hall families” (e.g. [24, 37]) or sometimes even simply “Hall bases” (e.g. [23], in which the generalized ones are called “Hall-Viennot bases”) or “Hall sets” (e.g. Bourbaki's choice in [4], anterior to Viennot's work).*

*We follow Viennot's definition (which seems to be the modern convention, also used in [29, 32]) and use the name “Hall set” to refer to Definition 1.26. To avoid confusion, we introduce the non-standard but more descriptive name “length-compatible Hall sets” instead of relying on the subtle distinction between “Hall sets” and “Philip Hall sets” (moreover, historically speaking, both Philip's and Marshall's definitions involved length-compatibility).*

Both in Philip Hall's description of his “commutator collecting process” and in Marshall Hall's definition of his “standard monomials”, it is explicit that the order between elements of the same length can be chosen arbitrarily.

Nevertheless, it seems that the mathematical literature and the software community has somehow progressively defined the most natural “Philip Hall basis of  $\mathcal{L}(X)$ ” as the Hall set on  $X$  whose order is given by the lexicographic order on the triple  $(|b|, \lambda(b), \mu(b))$  (which is indeed well-defined by induction on the length). This order appears explicitly in [28] associated with the name “basic commutator”, in [40] with the name “natural ordering” or in [15] with the name “optimal Philip Hall basis” (this work indeed claims that this particular choice allows to write an algorithm generating the basis up to some fixed length while minimizing the number of required bracket comparisons). This precise choice of length-compatible order also appears implicitly in many Lie algebraic packages, where it is linked with implementation choices (see e.g. `Hall.generate_hall_set` in SageMath [44] or `hall_basis.growup` in CoRoPa's LibAlgebra [5]). Some packages however make other choices (see e.g. `phb` in LTP [46]).

## 6.2 Preliminary results

In this paragraph,  $\mathcal{B} \subset \text{Br}(X)$  denotes a length-compatible Hall set.

**Lemma 6.3.** *Let  $a < b \in \mathcal{B}$  such that  $|b| < 2|a|$ . Then  $(a, b) \in \mathcal{B}$ .*

*Proof.* If  $b \in X$ , then  $(a, b) \in \mathcal{B}$ . Otherwise, we can write  $b = (a_1, b_1)$  with  $a_1 < b_1 \in \mathcal{B}$ . In particular, using (6.1),  $|a_1| \leq |b_1|$  thus  $|a_1| \leq \frac{1}{2}|b| < |a|$  which shows that  $a_1 < a$  and  $(a, b) \in \mathcal{B}$ .  $\square$

**Lemma 6.4.** *Let  $b = (a_1, b_1) \in \mathcal{B}$ ,  $a \in \mathcal{B}$  be such that  $a < a_1$  and  $h \in \text{supp}[a, a_1]$ . Then, either  $(h, b_1) \in \mathcal{B}$ , or  $(b_1, h) \in \mathcal{B}$  or  $h = b_1$ . In each case  $[h, b_1] \in \pm\mathcal{E}(\mathcal{B}) \cup \{0\}$ .*

*Proof.* We assume  $h \neq b_1$ .

*First case:  $h < b_1$ .* If  $b_1 \in X$ , then  $(h, b_1) \in \mathcal{B}$ . Otherwise, we may write  $b_1 = (a_2, b_2)$  with  $a_2 < b_2 \in \mathcal{B}$  and  $a_2 \leq a_1$ . Then  $|h| = |a| + |a_1| > |a_2|$ , so  $h > a_2$  and  $(h, b_1) \in \mathcal{B}$ .

*Second case:  $b_1 < h$ .* We have  $|h| = |a| + |a_1| \leq 2|a_1| \leq 2|b_1|$ . If this inequality is strict, then  $(b_1, h) \in \mathcal{B}$  by Lemma 6.3. If equality holds, then  $|a| = |a_1|$ , so  $(a, a_1) \in \mathcal{B}$  and  $h$  must be  $(a, a_1)$ . In this case, since  $b_1 > a_1 > a$ , we have again  $(b_1, h) \in \mathcal{B}$ .  $\square$

**Proposition 6.5.** *Let  $i \in \mathbb{N}^*$ ,  $b = (a_1, (\dots, (a_i, b_i) \dots)) \in \mathcal{B}$  and  $a \in \mathcal{B}$  such that  $a < a_i$ . Then*

$$[a_1, [\dots, [[a, a_i], b_i] \dots]] = \sum \alpha_c \mathcal{E}(c) \quad (6.2)$$

where the sum is finite,  $\alpha_c \in \mathbb{Z}$ ,  $\sum |\alpha_c| \leq 2^{i-1} \|[a, a_i]\|_{\mathcal{B}}$  and the  $c \in \mathcal{B}$  satisfy either  $\lambda(c) = a_1$  or  $c = (c', c'')$  with  $c' \in X$  or  $\lambda(c') \leq a_1$  and  $c'' \in X$  or  $\lambda(c'') \leq a_1$ .

*Proof.* The proof is by induction on  $i \in \mathbb{N}^*$ .

*Initialization for  $i = 1$ .* Let  $b = (a_1, b_1) \in \mathcal{B}$  and  $a \in \mathcal{B}$  such that  $a < a_1$ . We introduce the decomposition of  $[a, a_1]$  on  $\mathcal{B}$ :  $[a, a_1] = \sum_{h \in \mathcal{B}} \beta_h \mathcal{E}(h)$  with  $\beta_h \in \mathbb{K}$ . Then

$$[[a, a_1], b_1] = \sum_{h \in \mathcal{B}} \beta_h [h, b_1]. \quad (6.3)$$

By Lemma 6.4, for every  $h \in \mathcal{B}$  such that  $\beta_h \neq 0$ , then  $[h, b_1] \in \pm\mathcal{E}(\mathcal{B}) \cup \{0\}$ . Thus the above expression is the expansion of  $[[a, a_1], b_1]$  on  $\mathcal{B}$ , up to the sign of the basis elements. Therefore,

$$\|[a, a_1], b_1\|_{\mathcal{B}} \leq \sum_{h \in \mathcal{B}} |\beta_h| = \|[a, a_1]\|_{\mathcal{B}}. \quad (6.4)$$

Moreover, let us prove the desired structural properties on the supporting basis elements.



- Since  $b \in \mathcal{B}$ ,  $b_1 \in X$  or  $\lambda(b_1) \leq a_1$ .
- If  $|a| = |a_1|$ , then  $(a, a_1) \in \mathcal{B}$  so  $h = (a, a_1)$  and  $\lambda(h) = a < a_1$ . Otherwise,  $|h| = |a| + |a_1| < 2|a_1|$  so  $|\lambda(h)| < |a_1|$  so  $\lambda(h) < a_1$ .

*Induction.* Let  $i \geq 2$ . We assume the statement holds until  $(i-1)$ . Let  $b = (a_1, (\dots (a_i, b_i) \dots)) \in \mathcal{B}$  and  $a \in \mathcal{B}$  such that  $a < a_i$ . Then  $a_1 \geq \dots \geq a_i > a$  thus  $|a_1| \geq \dots \geq |a_i| \geq |a|$ . By the induction hypothesis,

$$[a_2, [\dots, [[a, a_i], b_i] \dots]] = \sum \alpha_c E(c), \quad (6.5)$$

where the sum is finite,  $\alpha_c \in \mathbb{Z}$ ,  $\sum |\alpha_c| \leq 2^{i-2} \|[a, a_i]\|_{\mathcal{B}}$  and the  $c \in \mathcal{B}$  satisfy either  $\lambda(c) = a_2$  or  $c = (c', c'')$  with  $c' \in X$  or  $\lambda(c') \leq a_2$  and  $c'' \in X$  or  $\lambda(c'') \leq a_2$ . Then,

$$[a_1, [a_2, [\dots, [[a, a_i], b_i] \dots]]] = \sum \alpha_c [a_1, c] \quad (6.6)$$

and, by the triangular inequality,

$$\|[a_1, [a_2, [\dots, [[a, a_i], b_i] \dots]]]\|_{\mathcal{B}} \leq \sum |\alpha_c| \|[a_1, c]\|_{\mathcal{B}}. \quad (6.7)$$

To conclude, it is thus sufficient to prove that, for every  $c \in \mathcal{B}$  such that  $\alpha_c \neq 0$ , then  $\|[a_1, c]\|_{\mathcal{B}} \leq 2$  and the expansion of  $[a_1, c]$  on  $\mathcal{B}$  only involves brackets of the desired form.

First, for each such  $c$ , since  $|c| = |a| + |b| - |a_1| > |b| - |a_1| \geq |a_1|$  (because  $b = (a_1, \mu(b)) \in \mathcal{B}$ ) one has  $a_1 < c$ . We must now consider three cases.

- If  $\lambda(c) = a_2$ , then  $(a_1, c) \in \mathcal{B}$  directly because  $a_2 \leq a_1$ .
- If  $\lambda(c) \leq a_1$ , then  $(a_1, c) \in \mathcal{B}$  directly also.
- If  $c = (c', c'')$  with  $a_1 < c' < c''$  and  $(c' \in X$  or  $\lambda(c') \leq a_2)$  and  $(c'' \in X$  or  $\lambda(c'') \leq a_2)$ , we use Jacobi's identity to write  $[a_1, c] = [[a_1, c'], c''] + [c', [a_1, c'']]$ . Using the pieces of information in the previous sentence suffices to ensure that these both brackets, up to sign, either vanish or belong to the basis and are of the desired form.

In particular, in all cases  $\|[a_1, c]\|_{\mathcal{B}} \leq 2$ , which concludes the proof.  $\square$

**Remark 6.6.** *The same proof also yields more information on the structure of elements  $c \in \mathcal{B}$  involved in (6.2). Indeed, they are of the form*

$$c = \left( a_1, \left( \dots, \left( a_p, \left( (a_{i_1}, (\dots, (a_{i_r}, v))) \quad , \quad (a_{j_1}, (\dots, (a_{j_s}, w))) \right) \dots \right) \right) \right) \quad (6.8)$$

where  $[v, w] = [h, b_i]$  for some  $h \in \mathcal{B}$  in the  $\text{supp}[a, a_i]$ ,  $p \in \llbracket 0, i-1 \rrbracket$  and the sets  $\{i_1 < \dots < i_r\}$ ,  $\{j_1 < \dots < j_s\}$  are a partition of  $\llbracket p+1, i-1 \rrbracket$ .

### 6.3 Proof of the main upper bound

We now prove the following refined version of Theorem 1.5.

**Theorem 6.7.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be a length-compatible Hall set. For all  $a < b \in \mathcal{B}$ ,*

$$\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta_a(b)-1} \leq 2^{\lfloor \frac{|b|}{|a|} \rfloor - 1}. \quad (6.9)$$

*Proof.* The second estimate is a consequence of the first one. Indeed, any leaf  $c$  of  $T_a(b)$  satisfies  $|c| \geq |a|$  because  $(a, c) \in \mathcal{B}$ , and these leaves are in number  $\theta_a(b)$ , thus  $|b| \geq \theta_a(b)|a|$ . Since  $\theta_a(b)$  is an integer, we can take the integral part of  $|b|/|a|$ .

We prove the first estimate by induction on  $q := \theta_a(b) \in \mathbb{N}^*$ .

The initialization for  $q = 1$  is given by Lemma 6.3.

Let  $q \geq 2$ . We assume the property proved until  $(q - 1)$ . Let  $a < b \in \mathcal{B}$  be such that  $\theta_a(b) = q$ . We define two sequences  $(a_i)_{i \in \llbracket 1, i_* \rrbracket}$ ,  $(b_i)_{i \in \llbracket 1, i_* \rrbracket}$  of  $\mathcal{B}$  by the following relations: for every  $i \in \llbracket 1, i_* \rrbracket$ ,

$$b = (a_1, (\dots, (a_i, b_i) \dots)), \quad (6.10)$$

i.e.  $a_i := \lambda(\mu^{i-1}(b))$ ,  $b_i := \mu^i(b)$ . The integer  $i_*$  is the smallest value of  $i$  for which  $b_i \in X$ . Then  $a_1 \geq \dots \geq a_{i_*}$  and  $a_i < b_i$  for every  $i \in \llbracket 1, i_* \rrbracket$ . We define

$$k := \begin{cases} \min\{i \in \llbracket 1, i_* - 1 \rrbracket; a_{i+1} < a\} & \text{if this set is not empty,} \\ i_* & \text{otherwise.} \end{cases} \quad (6.11)$$

Then  $(a, b_k) \in \mathcal{B}$ . Indeed, if  $k \in \llbracket 1, i_* - 1 \rrbracket$  then  $a_{k+1} < a \leq a_k < b_k$  so  $(a, b_k) = (a, (a_{k+1}, b_{k+1})) \in \mathcal{B}$ . If  $k = i_*$  then  $a < a_{i_*} < b_{i_*}$  and  $b_{i_*} \in X$  thus  $(a, b_{i_*}) \in \mathcal{B}$ .

Iterating Jacobi's identity, we get the expression

$$[a, b] = [a_1, [\dots, [a_k, [a, b_k]] \dots]] + \sum_{i=1}^k [a_1, [\dots, [[a, a_i], b_i] \dots]]. \quad (6.12)$$

The first term belongs to  $\mathbb{E}(\mathcal{B})$ . Indeed  $(a_k, (a, b_k)) \in \mathcal{B}$  because  $a_k < b_k < (a, b_k)$ . Moreover, for every  $i \in \llbracket 1, k - 1 \rrbracket$ ,  $a_i < b_i = (a_{i+1}(\dots, (a_k, b_k) \dots)) < (a_{i+1}, (\dots, (a_k, (a, b_k)) \dots))$  because the order is length-compatible and  $a_i \geq a_{i+1}$  thus  $(a_i, (\dots, ((a, a_k), b_k) \dots)) \in \mathcal{B}$ .

Then, by Proposition 6.5 and the induction assumption,

$$\|[a, b]\|_{\mathcal{B}} \leq 1 + \sum_{i=1}^k 2^{i-1} 2^{\theta_a(a_i)-1}. \quad (6.13)$$

Moreover, by definition of the map  $\theta_a$ , we have  $\sum_{i=1}^k \theta_a(a_i) = \theta_a(b) - \theta_a(b_k) \leq \theta_a(b) - 1$ . Finally, we get the conclusion by the following lemma.  $\square$

**Lemma 6.8.** *Let  $k \in \mathbb{N}^*$ ,  $q_1, \dots, q_k \in \mathbb{N}^*$  and  $Q := q_1 + \dots + q_k$ . Then*

$$\sum_{i=1}^k 2^{i-1} 2^{q_i-1} \leq 2^Q - 1. \quad (6.14)$$

*Proof.* The proof is by induction on  $k \in \mathbb{N}^*$ . If  $k = 1$ , then  $2^{q_1-1} \leq 2^{q_1} - 1 = 2^Q - 1$  holds. Let  $k \geq 2$ . We assume the result holds until  $(k - 1)$ . Let  $q_1, \dots, q_k$  and  $Q$  as in the statement. Then

$$\begin{aligned} \sum_{i=1}^k 2^{i-1} 2^{q_i-1} &= 2^{q_1-1} + 2 \sum_{i=1}^{k-1} 2^{i-1} 2^{q_{i+1}-1} \\ &\leq 2^{q_1-1} + 2(2^{Q-q_1} - 1) \\ &= 2^Q - 1 - (2^{q_1-1} - 1)(2^{Q-q_1+1} - 1) \leq 2^Q - 1, \end{aligned} \quad (6.15)$$

which concludes the proof and shows that equality holds if and only if  $q_i = 1$  for  $i \in \llbracket 1, k - 1 \rrbracket$ .  $\square$

**Remark 6.9.** *The same proof also yields more information on the structure of the elements of  $\text{supp}[a, b]$ : either  $(a_1, (\dots, (a_k, (a, b_k)) \dots))$  or of the form described in Remark 6.6 for  $i \in \llbracket 1, k \rrbracket$ . In any case, they satisfy the same structural property as in Proposition 6.5, i.e. either  $\lambda(c) = \lambda(b)$  or  $c = (c', c'')$  with  $(c' \in X \text{ or } \lambda(c') \leq \lambda(b))$  and  $(c'' \in X \text{ or } \lambda(c'') \leq \lambda(b))$ .*

## 6.4 Optimality cases

We now investigate the optimality of the estimate proved in the previous paragraph. We start with the following elementary bound on  $\theta_a(b)$ , which should be seen as a refinement of Proposition 4.7 in the case of length-compatible Hall sets when  $|X| = 2$ .

**Lemma 6.10.** *Let  $X = \{X_0, X_1\}$ . Let  $\mathcal{B} \subset \text{Br}(X)$  be a length-compatible Hall set such that  $X_0 < X_1$ . Then, for every  $a < b \in \mathcal{B}$  with  $|b| \geq 3$ ,*

$$\theta_a(b) \begin{cases} = |b| - 1 & \text{if } a = X_0 \text{ and } b = \text{ad}_{X_1}^n(X_0, X_1) \text{ for some } n \in \mathbb{N}, \\ \leq |b| - 2 & \text{otherwise.} \end{cases} \quad (6.16)$$

*Proof.* One already knows that  $\theta_a(b) \leq |b| - 1$  by the second item of Proposition 4.7. Let  $a < b \in \mathcal{B}$  with  $|b| \geq 3$  such that  $\theta_a(b) = |b| - 1$ . Then  $|b| - 1 = \theta_a(b) \leq \frac{|b|}{|a|}$ . Indeed, any leaf  $c$  of  $T_a(b)$  satisfies  $|c| \geq |a|$  because  $(a, c) \in \mathcal{B}$ , and these leaves are in number  $\theta_a(b)$ , thus  $|b| \geq \theta_a(b)|a|$ . Therefore,  $|a| \leq \frac{|b|}{|b| - 1} \leq \frac{3}{2}$  and  $a \in X$ . Necessarily,  $T_a(b)$  contains  $\theta_a(b) - 1 = |b| - 2$  leaves labeled by an indeterminate strictly greater than  $a$ , and one leaf labeled by the only basis element with length two, i.e.  $(X_0, X_1)$ . In conclusion,  $a = X_0$ , and  $b = \text{ad}_{X_1}^n(X_0, X_1)$  for some  $n \in \mathbb{N}$ , for which one checks that  $a < b \in \mathcal{B}$  and  $\theta_a(b) = n + 1 = |b| - 1$ .  $\square$

**Proposition 6.11.** *Let  $\mathcal{B} \subset \text{Br}(X)$  a length-compatible Hall set.*

1. *If  $|X| \geq 3$ , for every  $a < b \in \mathcal{B}$ ,  $\|[a, b]\|_{\mathcal{B}} \leq 2^{|b|-1}$ .*
2. *If  $X \supset \{X_0, X_1, X_2\}$  with  $X_0 < X_1 < X_2$ , for every  $n \in \mathbb{N}$ ,  $\text{ad}_{X_1}^n(X_2) \in \mathcal{B}$  and*

$$\|[X_0, \text{ad}_{X_1}^n(X_2)]\|_{\mathcal{B}} = 2^n \quad (6.17)$$

*so the previous estimate is optimal.*

3. *If  $|X| = 2$ , for every  $a < b \in \mathcal{B}$  with  $|b| \geq 3$ ,  $\|[a, b]\|_{\mathcal{B}} \leq 2^{|b|-3}$ .*
4. *If  $X = \{X_0, X_1\}$  with  $X_0 < X_1$ , for every  $n \in \mathbb{N}$ ,  $\text{ad}_{X_1}^n \text{ad}_{X_0}^2(X_1) \in \mathcal{B}$  and*

$$\|[X_0, \text{ad}_{X_1}^n \text{ad}_{X_0}^2(X_1)]\|_{\mathcal{B}} = 2^n \quad (6.18)$$

*so the previous estimate is optimal.*

*Proof.* We proceed step by step. Some computations are postponed to Section 5 since they can be carried out in more generality, with any Hall order.

1. This follows from Theorem 6.7 and the first item of Proposition 4.7.
2. This example is detailed in Theorem 5.2 in full generality.
3. Using Theorem 6.7 and Lemma 6.10, we obtain the expected bound on  $\|[a, b]\|_{\mathcal{B}}$ , except when  $a = X_0$  and  $b = \text{ad}_{X_1}^{n-1}(X_0, X_1)$  for some  $n \geq 2$ . Then, (5.12) yields  $\|[a, b]\|_{\mathcal{B}} \leq 2^{n-2} = 2^{|b|-3}$ .
4. For any  $j \in \mathbb{N}$ ,  $b_j := \text{ad}_{X_1}^j \text{ad}_{X_0}^2(X_1)$  belongs to  $\mathcal{B}$  because  $X_0 < X_1$  and the order is length-compatible. By iterating the Jacobi relation, we obtain

$$\begin{aligned} [X_0, \text{ad}_{X_1}^n \text{ad}_{X_0}^2(X_1)] &= \sum_{k=0}^{n-1} \text{ad}_{X_1}^k [[X_0, X_1], b_{n-k-1}] + \text{ad}_{X_1}^n \text{ad}_{X_0}^3(X_1) \\ &= \sum_{k=0}^{n-1} \sum_{j=0}^k \binom{k}{j} [\text{ad}_{X_1}^j [X_0, X_1], b_{n-j-1}] + \text{ad}_{X_1}^n \text{ad}_{X_0}^3(X_1) \\ &= \sum_{j=0}^{n-1} \alpha_j [\text{ad}_{X_1}^j [X_0, X_1], b_{n-j-1}] + \text{ad}_{X_1}^n \text{ad}_{X_0}^3(X_1) \end{aligned} \quad (6.19)$$

where  $\alpha_j := \sum_{k=j}^{n-1} \binom{k}{j}$ . Note that  $\text{ad}_{X_1}^n \text{ad}_{X_0}^3(X_1) \in \mathcal{B}$  because  $X_0 < X_1$ . Let us prove that  $\pm[\text{ad}_{X_1}^j[X_0, X_1], b_{n-j-1}] \in \mathbb{E}(\mathcal{B})$  for every  $j \in \llbracket 0, n-1 \rrbracket$ .

If  $\text{ad}_{X_1}^j(X_0, X_1) < b_{n-j-1}$  then  $(\text{ad}_{X_1}^j(X_0, X_1), b_{n-j-1}) \in \mathcal{B}$  because  $\lambda(b_{n-j-1})$  is either  $X_1$  or  $X_0$  thus  $\lambda(b_{n-j-1}) \leq \text{ad}_{X_1}^j(X_0, X_1)$  because the order is length-compatible.

If  $b_{n-j-1} < \text{ad}_{X_1}^j(X_0, X_1)$  then  $(b_{n-j-1}, \text{ad}_{X_1}^j(X_0, X_1)) \in \mathcal{B}$  because  $\lambda(\text{ad}_{X_1}^j(X_0, X_1))$  is either  $X_1$  or  $X_0$  thus  $\leq b_{n-j-1}$  because the order is length-compatible.

Eventually,

$$\| [X_0, \text{ad}_{X_1}^n \text{ad}_{X_0}^2(X_1)] \|_{\mathcal{B}} = \sum_{j=0}^{n-1} \alpha_j + 1 = \sum_{k=0}^{n-1} 2^k + 1 = 2^n. \quad (6.20)$$

This concludes the proof of the optimality of the length-based estimates.  $\square$

Interpreted in terms of the symmetric quantity  $\beta_n(\mathcal{B})$  defined in (5.1), these examples yield the following consequences.

**Corollary 6.12.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be length-compatible Hall set. When  $|X| \geq 3$ ,  $\beta_n(\mathcal{B}) = 2^{n-2}$  for every  $n \geq 2$ . When  $|X| = 2$ ,  $\beta_n(\mathcal{B}) = \max\{1, 2^{n-4}\}$  for every  $n \geq 2$ .*

## 7 Lyndon basis

The main goal of this section is to prove Theorem 1.6, which yields a sharp bound of the structure constants relative to the Lyndon basis. We start with definitions and a short introduction to the Lyndon basis in Section 7.1, then prove a refined version of Theorem 1.6 in Section 7.2. Eventually, we investigate the optimality of this estimate in Section 7.3.

### 7.1 Definitions and preliminary remarks

In this section,  $X$  is totally ordered and  $X^*$  (recall Definition 1.15) is endowed with the induced lexicographic order. We study the case of a classical Hall basis of  $\mathcal{L}(X)$  indexed by Lyndon words in  $X^*$ , the Lyndon basis. This basis is sometimes referred to as the “*Chen-Fox-Lyndon basis*” due to the important related results proved in [10], or as the “*Shirshov basis*” due to the work [38]. As in [32], we choose the name “*Lyndon basis*” for brevity and to highlight the source work [27] where Lyndon introduced “*standard sequences*” (which are now named Lyndon words).

For further details on the combinatorics of Lyndon words and their relations with Hall sets, the reader can refer to [26, Chapter 5], [32, Section 5.1].

**Definition 7.1** (Length, prefixes and suffixes). *We use the following notions for elements of  $X^*$ :*

- *If  $u \in X^*$ , its length  $|u|$  is the length of the corresponding sequence.*
- *If  $u, v \in X^*$ , we say that  $u$  is a prefix of  $v$  if there exists  $w \in X^*$  such that  $v = uw$ .*
- *If  $u, v \in X^*$ , we say that  $u$  is a suffix of  $v$  if there exists  $w \in X^*$  such that  $v = wu$ .*

**Definition 7.2** (Lyndon word). *A word  $w \in X^*$  is a Lyndon word if either  $w \in X$ , or for all  $u, v \in X^*$  such that  $w = uv$ ,  $w < vu$ . Denote by  $\text{Lyn}(X)$  the set of Lyndon words in  $X^*$ .*

As a consequence of this definition, if  $u, v \in \text{Lyn}(X)$  are such that  $u < v$ , then  $uv < v$ . Every Lyndon word that is not a letter can be written as the concatenation of two shorter Lyndon words. Such a factorization is not unique in general, but we can single out one of them: the *standard factorization* of  $w$  is the factorization  $w = uv$  with  $u, v \in \text{Lyn}(X)$  such that  $u$  has maximal length. Standard factorizations allow us to recursively define a map  $\text{br}: \text{Lyn}(X) \rightarrow \text{Br}(X)$  by mapping

each letter to itself and a Lyndon word  $w$  to  $\text{br}(w) = (\text{br}(u), \text{br}(v))$ , where  $w = uv$  is the standard factorization of  $w$ . Endow  $\text{Br}(X)$  with the preorder given by the lexicographic order on the foliage of trees, that we will call the *Lyndon order* (this order is not a total order, but it does not matter as we only intend to use its restriction to  $\text{br Lyn}(X)$ , which is a total order).

**Definition 7.3** (Lyndon basis). *The subset  $\mathcal{B} = \text{br}(\text{Lyn}(X))$  is a Hall set (see [32, Theorem 5.1]). The associated basis of  $\mathcal{L}(X)$  is called the Lyndon basis.*

In this section, if  $u \in \text{Lyn}(X)$ , extending the convention for left and right factors of the elements of  $\text{Br}(X)$ , we will denote by  $\lambda(u)$  the maximal strict Lyndon prefix of  $u$ , and by  $\mu(u)$  the suffix of  $u$  such that  $u = \lambda(u)\mu(u)$ . If  $u \in X$ , we use the convention that  $\lambda(u)$  is the empty word  $\varepsilon$ .

Let us recall some useful properties of the lexicographic order:

- If  $m, m', m'' \in X^*$  and  $m' < m''$ , then  $mm' < mm''$ .
- If  $m, m', m'' \in X^*$  and  $m < m' < mm''$ , then  $m$  is a prefix of  $m'$ .

**Lemma 7.4.** *Let  $b_1, b_2, b_3 \in \mathcal{B}$ . Then:*

1.  $\lambda(b_1) < b_1 < \mu(b_1)$ ,
2. if  $b_1 < b_2$  and  $|b_1| > |b_2|$  and  $b_1 \in \Lambda(b_3)$ , then  $b_3 < b_2$ ,
3. if  $b_1 < b_2 < b_3$  and  $b_1 \in \Lambda(b_3)$ , then  $b_1 \in \Lambda(b_2)$ ,
4. if  $b_2 < b_3$ , then  $(b_1, b_2) < (b_1, b_3)$ .

*Proof.* These are straightforward consequences of (1.10), the previously mentioned properties of the lexicographic order, and specific properties of Lyndon words.  $\square$

## 7.2 Proof of the main upper bound

We now prove the following refined version of Theorem 1.6.

**Theorem 7.5.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be the Hall set of the Lyndon basis. For all  $a < b \in \mathcal{B}$ ,*

$$\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta_a(b)-1} \quad (7.1)$$

*and for each  $c \in \text{supp}[a, b]$ ,  $a \in \Lambda(c)$  (so, in particular,  $a < c$ ),  $\lambda(c) \leq \max\{a, \lambda(b)\}$ , and  $c < b$ .*

*Proof.* We proceed by induction on  $n := \theta_a(b) \geq 1$  and execute a refined version of the decomposition algorithm Section 2.1.

*Initialization for  $n = 1$ .* By definition of  $\theta_a(b) = 1$  implies that  $(a, b) \in \mathcal{B}$  so is its own decomposition. Let  $c := (a, b)$ . Then one has all the desired properties, since  $\lambda(c) = a$ ,  $a \in \Lambda(c)$ ,  $\lambda(c) \leq \max\{a, \lambda(b)\}$  and  $c < b$  by item (1) of Lemma 7.4. Moreover,  $\|[a, b]\|_{\mathcal{B}} = 1 = 2^{1-1}$ .

*Induction.* Assume that the result is proved up to some  $n \in \mathbb{N}^*$ . Let  $a < b \in \mathcal{B}$  such that  $\theta_a(b) = n + 1$ . In particular,  $b = (\lambda(b), \mu(b))$  with  $a < \lambda(b) < \mu(b)$  (otherwise  $\theta_a(b) = 1$ ). By Jacobi's identity,

$$[a, b] = [[a, \lambda(b)], \mu(b)] + [\lambda(b), [a, \mu(b)]] \quad (7.2)$$

and we now treat both terms separately.

- *Study of  $[[a, \lambda(b)], \mu(b)]$ .* Since  $\theta_a(\lambda(b)) \leq \theta_a(b) - 1 \leq n$ , the induction hypothesis yields

$$[a, \lambda(b)] = \sum \alpha_d \mathbb{E}(d) \quad (7.3)$$

with  $d \in \mathcal{B}$ ,  $a \in \Lambda(d)$ ,  $\lambda(d) \leq \max\{a, \lambda^2(b)\}$ ,  $d < \lambda(b)$  and  $\sum |\alpha_d| \leq 2^{\theta_a(\lambda(b))-1}$ .

For each  $d$  in this sum,  $a < d < \lambda(b) < \mu(b)$  and, by Lemma 4.8,  $\theta_d(\mu(b)) \leq \theta_a(\mu(b)) \leq n$ , so the induction hypothesis applied to  $[d, \mu(b)]$  yields  $[d, \mu(b)] = \sum \beta_c \mathbb{E}(c)$  where  $d \in \Lambda(c)$ ,  $\lambda(c) \leq \max\{d, \lambda(\mu(b))\}$ ,  $c < \mu(b)$  and  $\sum |\beta_c| \leq 2^{\theta_a(\mu(b))-1}$ .

Since  $a \in \Lambda(d)$  and  $d \in \Lambda(c)$ ,  $a \in \Lambda(c)$ . Since  $b \in \mathcal{B}$ ,  $\lambda(\mu(b)) \leq \lambda(b)$  and, since  $d < \lambda(b)$ ,  $\lambda(c) \leq \max\{d, \lambda(\mu(b))\}$  implies that  $\lambda(c) \leq \lambda(b) = \max\{a, \lambda(b)\}$ . Moreover, as  $d < \lambda(b)$ ,  $|d| > |\lambda(b)|$  and  $d \in \Lambda(c)$ , item (2) of Lemma 7.4 implies that  $c < \lambda(b) < b$ .

This proves that the support of  $[[a, \lambda(b)], \mu(b)]$  has the desired properties and the size estimate  $\|[[a, \lambda(b)], \mu(b)]\|_{\mathcal{B}} \leq 2^{\theta_a(\lambda(b))-1} 2^{\theta_a(\mu(b))-1} = 2^{\theta_a(b)-2}$ .

- *Study of  $[\lambda(b), [a, \mu(b)]]$ .* Since  $\theta_a(\mu(b)) \leq n$ , the induction hypothesis yields

$$[a, \mu(b)] = \sum \alpha_d \mathbb{E}(d) \quad (7.4)$$

with  $d \in \mathcal{B}$ ,  $a \in \Lambda(d)$ ,  $\lambda(d) \leq \max\{a, \lambda(\mu(b))\}$ ,  $d < \mu(b)$  and  $\sum |\alpha_d| \leq 2^{\theta_a(\mu(b))-1}$ .

Let  $d \in \mathcal{B}$  be part of this sum. Unlike the previous case, we do not know how  $d$  and  $\lambda(b)$  compare, so we treat both cases separately.

- If  $d < \lambda(b)$ , since by Lemma 4.8,  $\theta_d(\lambda(b)) \leq \theta_a(\lambda(b)) \leq n$ , the induction hypothesis applied to  $[d, \lambda(b)]$  yields  $[d, \lambda(b)] = \sum \beta_c \mathbb{E}(c)$  where  $d \in \Lambda(c)$ ,  $\lambda(c) \leq \max\{d, \lambda^2(b)\}$ ,  $c < \lambda(b)$  and  $\sum |\beta_c| \leq 2^{\theta_a(\lambda(b))-1}$ . Since  $a \in \Lambda(d)$  and  $d \in \Lambda(c)$ ,  $a \in \Lambda(c)$ . Also,  $\lambda(c) \leq \max\{d, \lambda^2(b)\} < \lambda(b) = \max\{a, \lambda(b)\}$  and  $c < \lambda(b) < b$ . So  $c$  has the required properties and  $\|[[d, \lambda(b)]]\|_{\mathcal{B}} \leq 2^{\theta_a(\lambda(b))-1}$ .
- If  $\lambda(b) < d$ , since  $\lambda(d) \leq \max\{a, \lambda(\mu(b))\} \leq \lambda(b)$ , we have  $c := (\lambda(b), d) \in \mathcal{B}$ . Since  $a < \lambda(b) < d$  and  $a \in \Lambda(d)$ , item (3) of Lemma 7.4 yields  $a \in \Lambda(\lambda(b)) \subset \Lambda(c)$ . Obviously  $\lambda(c) = \lambda(b) \leq \max\{a, \lambda(b)\} = \lambda(b)$ . Finally, since  $d < \mu(b)$ , item (4) of Lemma 7.4 yields  $c = (\lambda(b), d) < (\lambda(b), \mu(b)) = b$ . So  $c$  has the required properties and  $\|[\lambda(b), d]\|_{\mathcal{B}} = 1$ .

This proves that the support of  $[[a, \lambda(b)], \mu(b)]$  has the desired properties and the size estimate  $\|[[a, \lambda(b)], \mu(b)]\|_{\mathcal{B}} \leq 2^{\theta_a(\mu(b))-1} \max\{2^{\theta_a(\lambda(b))-1}, 1\} = 2^{\theta_a(b)-2}$ .

Combining both studies and summing the estimates concludes the proof. □

### 7.3 Optimality cases

Theorem 7.5 yields the following optimal length-based estimates.

**Proposition 7.6.** *Let  $\mathcal{B} \subset \text{Br}(X)$  the Hall set of the Lyndon basis.*

1. If  $|X| \geq 3$ , for every  $a < b \in \mathcal{B}$ ,  $\| [a, b] \|_{\mathcal{B}} \leq 2^{|b|-1}$ .
2. If  $X \supset \{X_0, X_1, X_2\}$  with  $X_0 < X_1 < X_2$ , for every  $n \in \mathbb{N}$ ,  $\text{ad}_{X_1}^n(X_2) \in \mathcal{B}$  and

$$\| [X_0, \text{ad}_{X_1}^n(X_2)] \|_{\mathcal{B}} = 2^n \quad (7.5)$$

so the previous estimate is optimal.

3. If  $|X| = 2$ , for every  $a < b \in \mathcal{B}$  with  $|b| \geq 2$ ,  $\| [a, b] \|_{\mathcal{B}} \leq 2^{|b|-2}$ .
4. If  $X = \{X_0, X_1\}$  with  $X_0 < X_1$ , for every  $n \in \mathbb{N}^*$ ,  $\text{ad}_{X_0}^2(X_1) < \underline{\text{ad}}_{X_1}^n(X_0) \in \mathcal{B}$  and

$$\| [\text{ad}_{X_0}^2(X_1), \underline{\text{ad}}_{X_1}^n(X_0)] \|_{\mathcal{B}} = 2^{n-1} \quad (7.6)$$

so the previous estimate is optimal.

*Proof.* We proceed step by step. Some computations are postponed to Section 5 since they can be carried out in more generality, with any Hall order.

1. This follows from Theorem 7.5 and the first item of Proposition 4.7.
2. This example is detailed in Theorem 5.2 in full generality.
3. This follows from Theorem 7.5 and the second item of Proposition 4.7.
4. Let  $a := \text{ad}_{X_0}^2(X_1)$ . By Lemma B.2 (applied to the derivation  $\underline{\text{ad}}_{X_1}$  on  $\mathcal{L}(X)$ ,  $\nu \leftarrow n-1$ ,  $b_1 \leftarrow a$  and  $b_2 \leftarrow [X_0, X_1]$ ),

$$\begin{aligned} [a, \underline{\text{ad}}_{X_1}^n(X_0)] &= [a, \underline{\text{ad}}_{X_1}^{n-1}([X_0, X_1])] \\ &= \sum_{i=0}^{n-1} (-1)^i \binom{n-1}{i} \underline{\text{ad}}_{X_1}^{n-1-i}[\underline{\text{ad}}_{X_1}^i(a), [X_0, X_1]]. \end{aligned} \quad (7.7)$$

One checks that the elements of the right-hand side are (evaluations of) distinct basis elements, so the norm estimate readily follows.

This concludes the proof of the optimality of the Lyndon basis estimates.  $\square$

Interpreted in terms of the symmetric quantity  $\beta_n(\mathcal{B})$  defined in (5.1), these examples yield the following consequences.

**Corollary 7.7.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be the Hall set of the Lyndon basis over  $X$  with  $|X| \geq 3$ . Then, for every  $n \geq 2$ ,  $\beta_n(\mathcal{B}) = 2^{n-2}$ .*

**Proposition 7.8.** *Let  $\mathcal{B} \subset \text{Br}(X)$  be the Hall set of the Lyndon basis over  $X$  with  $|X| = 2$ . Then, for every  $n \geq 2$ ,  $\beta_n(\mathcal{B}) = \max\{1, F_{n-2}, 2^{n-5}\}$ .*

*Proof.* Let  $n \in \mathbb{N}$ . Let  $a < b \in \mathcal{B}$  with  $|a| + |b| = n$ . We separate cases depending on  $|a|$ . For  $m \in \mathbb{N}$ , we introduce the bracket  $A_m := \underline{\text{ad}}_{X_1}^m(X_0) \in \mathcal{B}$ .

*Case  $|a| \geq 3$ .* By the third item of Proposition 7.6,  $\|[a, b]\|_{\mathcal{B}} \leq 2^{|b|-2} \leq 2^{n-5}$ .

*Case  $|a| = 2$ .* Then  $a = (X_0, X_1)$ . By Proposition 4.7,  $\theta_a(b) \leq |b| - 1$ .

- If  $\theta_a(b) = |b| - 1$ , then  $b = \underline{\text{ad}}_{X_1}^m(X_0, X_1) = A_{m+1}$  with  $m = n - 4 \geq 1$ . Formula (5.6) (applied with  $a \leftarrow X_1$ ,  $b \leftarrow (X_0, X_1)$ ,  $n \leftarrow m$ ) yields

$$[A_1, A_{m+1}] = \sum_{p=0}^{\lfloor \frac{m-1}{2} \rfloor} (-1)^p \binom{m-1-p}{p} \underline{\text{ad}}_{X_1}^{m-1-2p}[A_{p+1}, A_{p+2}]. \quad (7.8)$$

This is indeed the decomposition of  $[a, b]$  on the basis, so  $\|[a, b]\|_{\mathcal{B}} = F_m = F_{n-4}$ .

- Otherwise,  $\theta_a(b) \leq |b| - 2$ , so, by Theorem 7.5,  $\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta_a(b)-1} \leq 2^{|b|-3} = 2^{n-5}$ .

*Case  $|a| = 1$ .* Then  $a = X_0$ . By Proposition 4.7,  $\theta_a(b) \leq |b| - 1$ .

- If  $\theta_a(b) = |b| - 1$ , then  $b = \underline{\text{ad}}_{X_1}^m(X_0) = A_m$  with  $m = n - 2 \geq 1$ . Formula (5.6) (applied with  $a \leftarrow X_1$ ,  $b \leftarrow X_0$ ,  $n \leftarrow m$ ) yields

$$[X_0, A_m] = \sum_{p=0}^{\lfloor \frac{m-1}{2} \rfloor} (-1)^p \binom{m-1-p}{p} \underline{\text{ad}}_{X_1}^{m-1-2p}[A_p, A_{p+1}]. \quad (7.9)$$

This is indeed the decomposition of  $[a, b]$  on the basis, so  $\|[a, b]\|_{\mathcal{B}} = F_m = F_{n-2}$ .

- If  $\theta_a(b) = |b| - 2$ ,  $b$  is of one of the following forms.

– Subcase  $b = \underline{\text{ad}}_{X_1}^m \text{ad}_{X_0}^2(X_1) =: S_m$  with  $m = n - 4 \geq 0$ . One has  $[X_0, S_0] = \text{ad}_{X_0}^3(X_1) \in \mathcal{B}$  so  $\|[X_0, S_0]\|_{\mathcal{B}} = 1$ . Moreover, by Jacobi's identity,

$$[X_0, S_{m+1}] = [X_0, [S_m, X_1]] = [[X_0, S_m], X_1] + [S_m, [X_0, X_1]]. \quad (7.10)$$

Since  $(S_m, (X_0, X_1)) \in \mathcal{B}$ , a straightforward induction on  $m$  proves that  $\|[X_0, S_m]\|_{\mathcal{B}} \leq m + 1$ . Hence  $\|[a, b]\|_{\mathcal{B}} \leq n - 3 \leq F_{n-2}$ .

– Subcase  $b = \underline{\text{ad}}_{X_1}^q(H_p)$  with  $p \geq 1$ ,  $q \in \mathbb{N}$ ,  $2p + q + 4 = n$ , where  $H_p := (A_p, A_{p+1})$ . We start with the case  $q = 0$ . Using Jacobi's identity,

$$[X_0, H_p] = [[X_0, A_p], A_{p+1}] - [[X_0, A_{p+1}], A_p]. \quad (7.11)$$

If  $d \in \text{supp}[X_0, A_p]$  (respectively  $d \in \text{supp}[X_0, A_{p+1}]$ ), then, by (7.9),  $d < A_{p+1}$  (resp.  $d < A_p$ ) and  $\theta_d(A_{p+1}) \leq |A_{p+1}| - 1 = p + 1$  (resp.  $\theta_d(A_p) \leq p$ ). Hence, by Theorem 7.5,

$$\|[X_0, H_p]\|_{\mathcal{B}} \leq F_p 2^p + F_{p+1} 2^{p-1} \leq 2^{p-1} F_{p+3}, \quad (7.12)$$

by (D.1). We now proceed by induction on  $q \in \mathbb{N}$ . By Jacobi's identity,

$$[X_0, \underline{\text{ad}}_{X_1}^{q+1}(H_p)] = [[X_0, \underline{\text{ad}}_{X_1}^q(H_p)], X_1] - [[X_0, X_1], \underline{\text{ad}}_{X_1}^q(H_p)]. \quad (7.13)$$

Since  $\theta_{(X_0, X_1)}(\underline{\text{ad}}_{X_1}^q(H_p)) = 2p - 1 + q$ , by Theorem 7.5 and induction,

$$\|[X_0, \underline{\text{ad}}_{X_1}^q(H_p)]\|_{\mathcal{B}} \leq \|[X_0, H_p]\|_{\mathcal{B}} + \sum_{r=0}^{q-1} 2^{(2p-1+r)-1} \leq 2^{p-1} F_{p+3} + 2^{2p-2} (2^q - 1) \quad (7.14)$$

using (7.12). Hence, for  $n$  large enough,  $\|[X_0, \underline{\text{ad}}_{X_1}^q(H_p)]\|_{\mathcal{B}} < 2^{n-5}$ . For small values of  $p$  and  $q$ , one checks that the right-hand side is indeed bounded by  $\max\{2^{n-5}, F_{n-2}\}$ .

- Otherwise,  $\theta_a(b) \leq |b| - 3$ , so, by Theorem 7.5,  $\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta_a(b)-1} \leq 2^{|b|-4} = 2^{n-5}$ .

This concludes the proof of the upper bounds. The lower bounds come from the fourth item of Proposition 7.6 (see (7.6)) and from the case  $\|[X_0, A_{n-2}]\|_{\mathcal{B}} = F_{n-2}$  (see (7.9)).  $\square$

## 8 A minimal Hall set

The main goal of this section is to prove Theorem 1.7, which illustrates the optimality of the lower bound of Theorem 5.5 when  $|X| = 2$  by exhibiting a Hall set  $\mathcal{B} \subset \text{Br}(X)$  with  $|X| = 2$  for which the growth of the structure constants has the magnitude of the Fibonacci sequence, i.e. such that  $\beta_{n+2}(\mathcal{B}) = F_n$ . We define a quite natural Hall set to answer this question, which seems new.

We give some motivations for the properties that we are looking for in Section 8.1. We define the Hall set we will consider in Section 8.2 and prove elementary structural properties in Section 8.3. We then start by a  $\theta$ -based size estimate for a particular family of right-nested brackets in Section 8.5. In Section 8.6, we then deduce from it a general size estimate, valid for all brackets, which distinguishes the role of the maximal indeterminate and allows to prove a slightly weaker version of Theorem 1.7 (see Corollary 8.19). In Section 8.7, we prove a refined version of our general estimate which concludes the proof of Theorem 1.7.

Eventually, in Section 8.8, we investigate an independent question concerning the optimal  $\theta$ -based estimate for our Hall set, which, quite surprisingly, turns out to be larger than the  $\theta$ -based estimates for length-compatible Hall sets and the Lyndon basis.

As in Section 5.3, throughout all this section,  $(F_\nu)_{\nu \in \mathbb{N}}$  denote the 0-based Fibonacci numbers.



## 8.1 Motivation

Let  $X := \{X_0, X_1\}$  and let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set with  $X_0 < X_1$ . We start by deriving necessary conditions for  $\beta_n(\mathcal{B})$  to have a geometric growth with a rate smaller than 2.

**Lemma 8.1.** *Let  $X := \{X_0, X_1\}$  and let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set with  $X_0 < X_1$ . With the notation of (5.1), assume that*

$$\lim_{n \rightarrow \infty} \frac{\beta_n(\mathcal{B})}{2^n} = 0. \quad (8.1)$$

*Then:*

- $\forall a \in \mathcal{B}$  such that  $a < X_1$ ,  $r(a, X_1) = +\infty$ ,
- $X_1 = \max \mathcal{B}$ ,
- $\forall a, b \in \mathcal{B} \setminus \{X_1\}$ ,  $\exists n \in \mathbb{N}$  such that  $\underline{\text{ad}}_{X_1}^n(a) > b$ .

*Proof.* We prove each item successively by contradiction.

*First item.* Assume that there exists  $a \in \mathcal{B}$  such that  $a < X_1$  and  $r(a, X_1) = r < +\infty$ . For  $n \geq r$ , let  $b_n := \text{ad}_{X_1}^{n-r}(\underline{\text{ad}}_{X_1}^r(a)) \in \mathcal{B}$  (see Lemma 4.6). Then the lower bound (5.11) of Theorem 5.5, combined with Proposition 3.5 (which is allowed since  $\{a, X_1\}$  is alphabetic and free) shows that  $2^{-n} \|[a, b_n]\|_{\mathcal{B}}$  does not go to zero as  $n \rightarrow +\infty$ , which contradicts (8.1) since  $|a| + |b_n| = 2|a| + n$ .

*Second item.* Assume that there exists  $b \in \mathcal{B}$  such that  $X_1 < b$ . We may choose  $b$  such that  $|b|$  is minimal (note that  $|b| \geq 2$  since  $X_0 < X_1$ ). In this case, by minimality of  $b$ ,  $\lambda(b) < \mu(b) \leq X_1$ . Let  $a := \lambda(b)$ . By hypothesis,  $a < X_1 < b$ , and since  $r(a, X_1) = +\infty$ , for all  $n \in \mathbb{N}$ ,  $\underline{\text{ad}}_{X_1}^n(a) < X_1 < b$ . This shows in particular that for all  $n \in \mathbb{N}$ ,  $(\underline{\text{ad}}_{X_1}^n(a), b) \in \mathcal{B}$ , because  $\lambda(b) = a < \underline{\text{ad}}_{X_1}^n(a)$ . Let  $b_n := \text{ad}_{X_1}^n(b) \in \mathcal{B}$ . By Lemma B.2:

$$[a, b_n] = \sum_{k=0}^n \binom{n}{k} \text{ad}_{X_1}^{n-k}([\underline{\text{ad}}_{X_1}^k(a), b]), \quad (8.2)$$

and the previous discussion shows that for all  $k, n \in \mathbb{N}$ ,  $\text{ad}_{X_1}^{n-k}([\underline{\text{ad}}_{X_1}^k(a), b]) \in \mathcal{E}(\mathcal{B})$ , so that  $\|[a, b_n]\|_{\mathcal{B}} = 2^n$ , with  $|a| + |b_n| = n + |a| + |b|$ , which contradicts (8.1).

*Third item.* Let  $a_0 \in \mathcal{B} \setminus \{X_1\}$  and assume that there exists  $b \in \mathcal{B}$  with  $b \neq X_1$  such that for all  $n \in \mathbb{N}$ ,  $\underline{\text{ad}}_{X_1}^n(a_0) < b$ . We may assume that  $|b|$  is minimal (by assumption,  $|b| \geq 2$ ). By minimality of  $b$  and since  $\lambda(b) \neq X_1$ , there exists  $k \geq 0$  such that  $\underline{\text{ad}}_{X_1}^k(a_0) \geq \lambda(b)$ . Let  $k \in \mathbb{N}$  be minimal such that  $\underline{\text{ad}}_{X_1}^k(a_0) \geq \lambda(b)$ , and let  $a := \underline{\text{ad}}_{X_1}^k(a_0)$ . Then, by construction, for all  $n \in \mathbb{N}$  we have  $\underline{\text{ad}}_{X_1}^n(a) < b$  and  $(\underline{\text{ad}}_{X_1}^n(a), b) \in \mathcal{B}$ . Therefore, if we let  $b_n := \underline{\text{ad}}_{X_1}^n(b)$ , a computation similar to that of the second item using Lemma B.2 shows that  $\|[a, b_n]\|_{\mathcal{B}} = 2^n$ , which is again a contradiction.  $\square$

This lemma shows that if we intend to exhibit a Hall set  $\mathcal{B}$  for which the growth of the structure constants has the magnitude of the Fibonacci sequence, then it is necessary that  $X_1$  be maximal in  $\mathcal{B}$  (which excludes length-compatible Hall sets), but also that the third property (which is reminiscent of the Archimedean property) of the lemma holds. For example, in the case of the Lyndon basis over  $\{X_0 < X_1\}$ , the element  $X_1$  is maximal but the third property does not hold (for example,  $\underline{\text{ad}}_{X_1}^n((X_0, (X_0, X_1))) < (X_0, X_1)$  for all  $n \in \mathbb{N}$ ).

An example of a known Hall set satisfying all three conditions is the Spitzer-Foata basis (as named and described in [48, Chapter I, Section 5]; see also [49]). The order defining this Hall set is compatible with the increasing order on the ratio  $s(b) := n_1(b)/n_0(b) \in [0; +\infty]$ , where  $n_i(b)$  denotes the number of occurrences of  $X_i$  in  $b$ . In particular,  $X_1$  is maximal and, for every  $a, b \in \mathcal{B} \setminus \{X_1\}$ ,  $s(\underline{\text{ad}}_{X_1}^n(a)) \rightarrow +\infty$  as  $n \rightarrow +\infty$ , so it is strictly greater than  $s(b)$  for  $n$  large enough, and the third property is satisfied. Unfortunately, numerical simulations we conducted indicate that this basis does not satisfy  $\beta_{n+2}(\mathcal{B}) = F_n$ , which motivated our construction of a new basis, explained in Section 8.2. We do not know whether the basis we construct is the unique one satisfying  $\beta_{n+2}(\mathcal{B}) = F_n$  or if others exist.

## 8.2 Definition of our minimal Hall set

We will base our construction on the following “ $\lambda, \mu$ ”-lexicographic order, where one indeterminate is considered maximal.

**Definition 8.2.** Let  $X = \{X_0, X_1\}$  and consider the following order on  $\text{Br}(X)$ :

- $X_0$  is minimal and  $X_1$  is maximal, i.e.,  $X_0 < \text{Br}(X) \setminus X < X_1$ ,
- for  $t_1, t_2 \in \text{Br}(X) \setminus X$ ,  $t_1 < t_2$  iff either  $\lambda(t_1) < \lambda(t_2)$ , or  $\lambda(t_1) = \lambda(t_2)$  and  $\mu(t_1) < \mu(t_2)$ .

**Lemma 8.3.** There exists a unique Hall set  $\mathcal{B} \subset \text{Br}(X)$  associated with the order of Definition 8.2.

*Proof.* This order is not a Hall order on  $\text{Br}(X)$  because there exists  $t \in \text{Br}(X)$  such that  $\lambda(t) > t$ , for instance  $t = (X_1, X_0)$ . But we can construct  $\mathcal{B}$  by applying Lemma 1.37 to the set  $G := \{b \in \text{Br}(X); X_1 \notin \Lambda(b)\} \cup \{X_1\}$  (see Definition 1.12). Clearly,  $G$  is a  $\lambda$ -stable subset of  $\text{Br}(X)$ , containing  $X$ , endowed with a total order and for every  $b_1 < b_2 \in G$ ,  $(b_1, b_2) \in G$ . It remains to prove that  $<$  is a Hall order on  $G$ . By contradiction, let  $b \in G \setminus X$  of minimal length such that  $\lambda(b) > b$ . Since  $\lambda(b) > b$ ,  $\lambda(b) \neq X_0$ . Since  $b \in G$ ,  $\lambda(b) \neq X_1$ . Hence  $\lambda(b) \in G \setminus X$  and satisfies  $\lambda(\lambda(b)) < \lambda(b)$  since  $b$  was of minimal length. By definition of the order, this implies  $\lambda(b) < b$ .  $\square$

In the sequel of this section,  $\mathcal{B}$  denotes the Hall set constructed in Lemma 8.3.

**Remark 8.4.** By definition,  $X_1 = \max \mathcal{B}$  (thus  $r(a, X_1) = +\infty$  for all  $a \in \mathcal{B}$ ), and if  $a, b \in \mathcal{B}$  with  $a, b \neq X_1$ , then a straightforward induction on  $|b|$  shows that there exists  $n \in \mathbb{N}$  such that  $\text{ad}_{X_1}^n(a) > b$ . Therefore, all the necessary conditions of Lemma 8.1 hold for  $\mathcal{B}$ .

## 8.3 Elementary structural properties

We state in the following lemmas useful properties of this Hall set.

**Lemma 8.5.** Let  $a < b \in \mathcal{B}$  with  $b \neq X_1$  and  $\lambda(b) < a$ . For every  $n < n' \in \mathbb{N}$  and  $p < p' \in \mathbb{N}$ ,

$$\text{ad}_a^n(b) < \text{ad}_a^{n'}(b) < \text{ad}_a^{p'}(X_1) < \text{ad}_a^p(X_1). \quad (8.3)$$

*Proof.* These inequalities stem from Definition 8.2. First, since  $\lambda(b) < a$ ,  $b < (a, \text{ad}_a^{n'-n-1}(b))$ , which yields the first inequality by applying  $\text{ad}_a^n$  to both sides. Second, since  $X_1$  is maximal,  $\text{ad}_a^{p'-p}(b)(X_1) < X_1$ , which yields the third inequality by applying  $\text{ad}_a^p$  to both sides. Eventually, when  $n' \geq p'$ , the middle inequality holds because  $\text{ad}_a^{n'-p'}(b) < X_1$  and, when  $n' < p'$ , it holds because  $b < \text{ad}_a^{p'-n'}(X_1)$  since  $\lambda(b) < a$  by assumption.  $\square$

**Lemma 8.6.** Let  $b \in \mathcal{B} \setminus X$ . Then  $b < \mu(b)$  iff  $b = \text{ad}_{\lambda(b)}^m(X_1)$  for some  $m \in \mathbb{N}^*$ .

*Proof.* Let  $b \in \mathcal{B} \setminus X$ . There exists  $m \in \mathbb{N}^*$  such that  $b = \text{ad}_{\lambda(b)}^m(v)$  with  $v = X_1$  or  $v \in \mathcal{B} \setminus X$  with  $\lambda(v) < \lambda(b)$ . Since  $\mu(b) = \text{ad}_{\lambda(b)}^{m-1}(v)$ , Lemma 8.5 proves that  $b < \mu(b)$  iff  $v = X_1$ .  $\square$

**Lemma 8.7.** Let  $a \in \mathcal{B}$  and  $b \in \mathcal{B} \setminus X$  with  $\mu(b) \neq X_1$  such that  $a < \mu(b)$ . Then  $a \leq b$ .

*Proof.* By contradiction, if  $b < a < \mu(b)$ , then  $\lambda(b) \leq \lambda(a) \leq \lambda(\mu(b)) \leq \lambda(b)$  so there exists  $m \in \mathbb{N}^*$  such that  $a = \text{ad}_{\lambda(b)}^m(v)$  with  $v = X_1$  or  $v \in \mathcal{B} \setminus X$  with  $\lambda(v) < \lambda(b)$ . Since  $b < \mu(b)$ , by Lemma 8.6,  $b = \text{ad}_{\lambda(b)}^n(X_1)$  for some  $n \in \mathbb{N}^*$ . Moreover,  $n \geq 2$  because  $\mu(b) \neq X_1$  by assumption. Using Lemma 8.5,  $b < a$  implies  $v = X_1$  and  $b < a < \mu(b)$  implies  $n > m > n - 1$ .  $\square$

**Lemma 8.8.** Let  $a < b \in \mathcal{B}$  with  $b \notin X$ . Then  $\theta_a(b) \leq 2$  if both of the following conditions hold

- either  $\lambda(b) = X_0$  or  $\lambda^2(b) \leq a$ ,
- either  $\mu(b) = X_1$  or  $\lambda(\mu(b)) \leq a$ .

*Proof.* If  $\lambda(b) \leq a$  then  $\theta_a(b) = 1$ . Otherwise  $a < \lambda(b) < \mu(b)$  so  $\theta_a(b) = \theta_a(\lambda(b)) + \theta_a(\mu(b))$ . The first (resp. second) condition gives  $\theta_a(\lambda(b)) = 1$  (resp.  $\theta_a(\mu(b)) = 1$ ).  $\square$

## 8.4 Structural properties of supporting basis elements

We prove the following structural properties on elements of the support of  $[a, b]$ . These properties are of independent interest for the understanding of  $\mathcal{B}$  and will play a key role in estimating the size of its structure constants, since they allow to stop the decomposition algorithm much earlier than in the general case.

**Theorem 8.9.** *For every  $a < b \in \mathcal{B}$ , each  $c \in \text{supp}[a, b]$  satisfies:  $c \notin X$ ,*

- $a \leq \lambda(c) < b$  where the first relation is an equality if and only if  $(a, b) \in \mathcal{B}$ ,
- and, when  $b \neq X_1$ , either  $c \leq \lambda(b)$  or  $\theta_{\lambda(b)}(c) \leq 2$ .

*Proof.* The relation between  $a$  and  $\lambda(c)$  is proved in Theorem 2.1 for any Hall set (see (2.2)). Thus, we only prove the other relations.

We proceed, as in Proposition 4.17, by induction on  $\theta_a(b)$ , by applying the classical rewriting scheme. We refer to the proof of Proposition 4.17 for a detailed justification of why the induction on  $\theta_a(b)$  is legitimate in this setting.

If  $\theta_a(b) = 1$ , then  $c = (a, b) \in \mathcal{B}$  and satisfies  $\lambda(c) = a < b$ . If  $b \neq X_1$ , then  $\lambda(b) \leq a < c$ . If  $a = X_0$ , then  $\theta_{\lambda(b)}(c) = 1$ . Otherwise, by Definition 8.2,  $\lambda(a) \leq \lambda(b)$ , so, by Lemma 8.8,  $\theta_{\lambda(b)}(c) \leq 2$ .

From now on, we consider  $\theta_a(b) > 1$  and we assume the result proved for strictly smaller values. Then  $b \neq X_1$ . Moreover,  $\lambda(b) \notin X$ , because  $X_0 \leq a < \lambda(b) < \mu(b) \leq X_1$ , thus  $\lambda^2(b)$  and  $\mu(\lambda(b))$  are well defined. By Jacobi's identity

$$[a, b] = [[a, \lambda(b)], \mu(b)] + [\lambda(b), [a, \mu(b)]]. \quad (8.4)$$

*A preliminary remark.* When  $\mu(b) \neq X_1$ , for each  $c \in \text{supp}[a, b]$ ,  $\lambda(c) < \mu(b)$  implies  $\lambda(c) < b$ . Indeed, by Lemma 8.7,  $\lambda(c) < \mu(b)$  implies  $\lambda(c) \leq b$  and, by Lemma 4.16, one cannot have  $\lambda(c) = b$ .

*Study of the first term.* By induction,

$$[a, \lambda(b)] = \sum \gamma_d \mathbb{E}(d), \quad \text{where } a \leq \lambda(d) < \lambda(b) \text{ and } (\text{either } d \leq \lambda^2(b) \text{ or } \theta_{\lambda^2(b)}(d) \leq 2). \quad (8.5)$$

Let  $d \in \text{supp}[a, \lambda(b)]$ . Then  $a < d < b$  because  $a \leq \lambda(d) < d$  and  $\lambda(d) < \lambda(b)$ .

- *Case  $\mu(b) = X_1$ .* Then  $c := (d, \mu(b)) = (d, X_1) \in \mathcal{B}$  satisfies  $\lambda(c) = d < b$ . If  $\lambda(b) < c$ , since  $\lambda(d) < \lambda(b)$ , then, by Lemma 8.8,  $\theta_{\lambda(b)}(c) \leq 2$ .
- *Case  $\mu(b) \neq X_1$ , so  $\mu(b) \in \mathcal{B} \setminus X$  and  $\lambda(\mu(b)) \leq \lambda(b)$ .*
  - *Subcase  $\mu(b) < d$ .* Then  $c := (\mu(b), d) \in \mathcal{B}$  because  $\lambda(d) < \lambda(b) < \mu(b)$ . Moreover,  $\lambda(c) = \mu(b) < d < b$ . Then  $\lambda(b) < \mu(b) < c$  and, since  $\lambda(\mu(b)) \leq \lambda(b)$  and  $\lambda(d) < \lambda(b)$ , then, by Lemma 8.8,  $\theta_{\lambda(b)}(c) \leq 2$ .
  - *Subcase  $d < \mu(b)$ .* By induction,  $[d, \mu(b)] = \sum \alpha_c \mathbb{E}(c)$  where  $\lambda(c) < \mu(b)$  and, either  $c \leq \lambda(\mu(b))$ , or  $\theta_{\lambda(\mu(b))}(c) \leq 2$ . By the preliminary remark,  $\lambda(c) < b$ . If  $\lambda(b) < c$ , by Lemma 4.8,  $\theta_{\lambda(b)}(c) \leq \theta_{\lambda(\mu(b))}(c) \leq 2$ .

*Study of the second term.*

- *Case  $\mu(b) = X_1$ .* Then  $d := (a, \mu(b)) = (a, X_1) \in \mathcal{B}$  and  $a < d$  by the Hall order property.
  - If  $d < \lambda(b)$ , by induction,  $[d, \lambda(b)] = \sum \alpha_c \mathbb{E}(c)$  where  $\lambda(c) < \lambda(b)$  and either  $c \leq \lambda^2(b)$  or  $\theta_{\lambda^2(b)}(c) \leq 2$ . Thus,  $\lambda(c) < b$  because  $\lambda(b) < b$ . If  $\lambda(b) < c$ , by Lemma 4.8,  $\theta_{\lambda(b)}(c) \leq \theta_{\lambda^2(b)}(c) \leq 2$ .

– If  $\lambda(b) < d$  then  $c := (\lambda(b), d) = (\lambda(b), (a, X_1)) \in \mathcal{B}$  because  $a < \lambda(b)$ . Moreover  $\lambda(c) = \lambda(b) < b$ ,  $\lambda(b) < c$  and  $\theta_{\lambda(b)}(c) = 1$ .

• *Case  $\mu(b) \neq X_1$ .* By induction,

$$[a, \mu(b)] = \sum \gamma_d \mathbf{E}(d) \text{ where } a \leq \lambda(d) < \mu(b) \text{ and (either } d \leq \lambda(\mu(b)) \text{ or } \theta_{\lambda(\mu(b))}(d) \leq 2). \quad (8.6)$$

Let  $d \in \text{supp}[a, \mu(b)]$ . Then  $a < d$  because  $a \leq \lambda(d) < d$ .

– *Subcase  $d < \lambda(b)$ .* By induction,  $[d, \lambda(b)] = \sum \alpha_c \mathbf{E}(c)$  where  $\lambda(c) < \lambda(b)$  and (either  $c \leq \lambda^2(b)$  or  $\theta_{\lambda^2(b)}(c) \leq 2$ ), which implies the expected relations, because the order is a Hall order on  $\mathcal{B}$  (so  $\lambda(b) < b$  and  $\lambda^2(b) < \lambda(b)$ ) and thanks to Lemma 4.8.

– *Subcase  $\lambda(b) < d$ .*

\* *Subsubcase  $\lambda(d) \leq \lambda(b)$ .* Then  $c := (\lambda(b), d) \in \mathcal{B}$  satisfies  $\lambda(c) = \lambda(b) < b$  by the Hall order property and  $\lambda(b) < c$  and  $\theta_{\lambda(b)}(c) = 1$ .

\* *Subsubcase  $\lambda(b) < d$ .* This implies  $\lambda(\mu(b)) < d$  and, by (8.6) and Lemma 4.8,  $\theta_{\lambda(b)}(d) \leq \theta_{\lambda(\mu(b))}(d) \leq 2$ . Thus, for each  $c \in \text{supp}[\lambda(b), d]$ ,  $\lambda(b) < \lambda(c)$  by (2.2), and  $\theta_{\lambda(b)}(c) = \theta_{\lambda(b)}(d) \leq 2$  by Lemma 4.16.

To get the conclusion, we only need to check that  $\lambda(c) < b$ . By Theorem 4.12,  $c \in \mathbf{i}(\text{Tr}[\lambda(b) < \lambda(d) < \mu(d)])$ . By the Hall order property,  $c_1 := \lambda(b) < b$ . By (8.6),  $c_2 := \lambda(d) < \mu(b)$ . Hence,  $c_3 := (\lambda(b), \lambda(d)) < b$  by Definition 8.2. Since  $\lambda(c) < \mu(c)$  and one of them is of one of the  $c_i$ 's, we obtain  $\lambda(c) < b$  or  $\lambda(c) < \mu(b)$ , hence  $\lambda(c) < b$  by the preliminary remark.  $\square$

## 8.5 A first estimate for right-nested trees

In this paragraph, we prove a bound on  $\|[a, b]\|_{\mathcal{B}}$  depending only on  $\theta_a(b)$  in the particular case where  $\lambda^2(b) \leq a$ . We tighten this bound and remove the hypothesis in Section 8.8. We start here this lighter version since it is sufficient for the proof of Theorem 1.7.

**Lemma 8.10.** *Let  $a < b \in \mathcal{B}$  such that  $(a, b) \notin \mathcal{B}$  and  $\lambda^2(b) \leq a$ . Then  $T_a(b)$  is right-nested, in the sense that it has the following form*

$$T_a(b) = \langle \lambda(b), \langle \lambda(\mu(b)), \langle \dots \langle \lambda(\mu^{\theta-2}(b)), \mu^{\theta-1}(b) \rangle \rangle \dots \rangle \quad (8.7)$$

where  $\theta = \theta_a(b)$  and, in particular, only the last leaf,  $\mu^{\theta-1}(b)$ , can be equal to  $X_1$ .

*Proof.* The assumption  $(a, b) \notin \mathcal{B}$  implies that  $b \neq X_1$ , so  $\lambda(b)$  makes sense and that  $\lambda(b) \neq X_0$ , so  $\lambda^2(b)$  is also well-defined. We proceed by induction on  $\theta_a(b)$ . When  $\theta_a(b) = 2$ , it is true in any Hall set that  $T_a(b) = \langle \lambda(b), \mu(b) \rangle$ . When  $\theta_a(b) \geq 3$ , since  $\lambda^2(b) \leq a < \lambda(b)$ , by definition,  $T_a(b) = \langle \lambda(b), T_a(\mu(b)) \rangle$ . The result follows by induction because  $\lambda^2(\mu(b)) \leq \lambda^2(b) \leq a$ . (The first inequality stems from Definition 8.2 and from the fact that  $\lambda(\mu(b)) \leq \lambda(b)$ , since  $b \in \mathcal{B}$ ).  $\square$

We start by a definition which allows to distinguish, within  $\|[a, b]\|_{\mathcal{B}}$ , the part of the norm relative to the different categories of supporting basis elements.

**Definition 8.11.** *Let  $a < b \in \mathcal{B}$  with  $b \neq X_1$ . Writing  $[a, b] = \sum \alpha_c \mathbf{E}(c)$  where the sum ranges over  $\text{supp}[a, b]$  and  $\alpha_c \in \mathbb{K}$ , we define, for  $i \in \llbracket 1, 2 \rrbracket$ ,*

$$N_i(a, b) := \sum_{\theta_{\lambda(b)}(c)=i} |\alpha_c|. \quad (8.8)$$

**Lemma 8.12.** *Let  $a < b \in \mathcal{B}$  such that  $(a, b) \notin \mathcal{B}$  and  $\lambda^2(b) \leq a$ . For each  $c \in \text{supp}[a, b]$ ,  $\lambda(b) < c$ . Moreover*

$$\|[a, b]\|_{\mathcal{B}} = N_1(a, b) + N_2(a, b). \quad (8.9)$$

*Proof.* Let  $c \in \text{supp}[a, b]$ . Since  $(a, b) \notin \mathcal{B}$ , by (2.2),  $a < \lambda(c)$ . Hence  $\lambda^2(b) < \lambda(c)$ , so  $\lambda(b) < c$  by Definition 8.2. Then (8.9) follows from Theorem 8.9 since the case  $c \leq \lambda(b)$  is excluded.  $\square$

**Lemma 8.13.** *Let  $a < b \in \mathcal{B}$  with  $\theta_a(b) = 2$ . Then  $\|[a, b]\|_{\mathcal{B}} \leq 2$  and  $N_2(a, b) \leq 1$ .*

*Proof.* Since  $\theta_a(b) = 2$ ,  $a < \lambda(b)$ ,  $(a, \lambda(b)) \in \mathcal{B}$  and  $(a, \mu(b)) \in \mathcal{B}$ . By Jacobi's identity

$$[a, b] = [(a, \lambda(b)), \mu(b)] + [\lambda(b), (a, \mu(b))]. \quad (8.10)$$

*First term.*

- If  $(a, \lambda(b)) < \mu(b)$  then  $c_1 := ((a, \lambda(b)), \mu(b)) \in \mathcal{B}$  because  $\lambda(\mu(b)) \leq a < (a, \lambda(b))$ .
- If  $\mu(b) < (a, \lambda(b))$  then  $c_1 := (\mu(b), (a, \lambda(b))) \in \mathcal{B}$  because  $a < \mu(b)$ .

*Second term.*

- If  $\lambda(b) < (a, \mu(b))$ , then  $c_2 := (\lambda(b), (a, \mu(b))) \in \mathcal{B}$  because  $a < \lambda(b)$ .
- If  $(a, \mu(b)) < \lambda(b)$ , then  $c_2 := ((a, \mu(b)), \lambda(b)) \in \mathcal{B}$  because  $\lambda^2(b) \leq a < (a, \mu(b))$ .

In particular,  $\theta_{\lambda(b)}(c_2) = 1$  because  $\lambda(c_2) = \min\{\lambda(b), (a, \mu(b))\} \leq \lambda(b)$ . So  $N_2(a, b) \leq 1$ .  $\square$

We start with an easy case where  $\lambda^2(b) < a$ , leading to a straightforward geometric bound.

**Proposition 8.14.** *Let  $a < b \in \mathcal{B}$  such that  $(a, b) \notin \mathcal{B}$  and  $\lambda^2(b) < a$ . Then*

$$\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta-1} \quad \text{and} \quad N_2(a, b) \leq 2^{\theta-1} - 1, \quad \text{where } \theta = \theta_a(b). \quad (8.11)$$

*Proof.* By Definition 8.2,  $\lambda^2(b) < a$  implies that  $\lambda(b) < (a, \lambda(b))$ . We proceed by induction on  $\theta_a(b) \geq 2$ . Initialization is proved by Lemma 8.13. We perform the inductive step. Assume  $\theta_a(b) \geq 3$ . Then  $\theta_a(\mu(b)) = \theta_a(b) - 1 \geq 2$  thus  $a < \lambda(\mu(b))$ ,  $\lambda(\mu(b)) \notin X$  and  $\lambda^2(\mu(b))$  is well defined. Using Jacobi's identity,

$$[a, b] = [(a, \lambda(b)), \mu(b)] + [\lambda(b), [a, \mu(b))]. \quad (8.12)$$

*First term of (8.12).* We have  $(a, \lambda(b)) < \mu(b)$  by Definition 8.2 because  $a < \lambda(\mu(b))$ . Then  $c_1 := ((a, \lambda(b)), \mu(b)) \in \mathcal{B}$  because, since  $b \in \mathcal{B}$ ,  $\lambda(\mu(b)) \leq \lambda(b) < (a, \lambda(b))$ . Moreover  $\theta_{\lambda(b)}(c_1) = 2$  because  $\lambda(b) < \lambda(c_1) = (a, \lambda(b))$ .

*Second term of (8.12).* The induction hypothesis applies to  $[a, \mu(b)]$  because  $a < \mu(b)$ ,  $\lambda^2(\mu(b)) \leq \lambda^2(b) < a$  by Definition 8.2 and assumption, and  $\theta_a(\mu(b)) = \theta_a(b) - 1 \geq 2$ . Hence  $[a, \mu(b)] = \sum \alpha_d \mathbf{E}(d)$  with the claimed size estimates. Let  $d \in \text{supp}[a, \mu(b)]$ .

- If  $\theta_{\lambda(b)}(d) = 1$ , then  $c := (\lambda(b), d) \in \mathcal{B}$  and  $\theta_{\lambda(b)}(c) = 1$ .
- If  $\theta_{\lambda(b)}(d) = 2$ , by Theorem 4.12,  $\|[\lambda(b), d]\|_{\mathcal{B}} \leq 2$  and, by Lemma 4.16, for each  $c \in \text{supp}[\lambda(b), d]$ ,  $\theta_{\lambda(b)}(c) = 2$ .

Moreover, by Lemma 4.8,  $\theta_{\lambda(b)}(d) \leq \theta_{\lambda(\mu(b))}(d)$ .

*Conclusion.* This analysis proves  $N_2(a, b) \leq 1 + 2N_2(a, \mu(b))$  and that

$$\|[a, b]\|_{\mathcal{B}} \leq 1 + N_1(a, \mu(b)) + 2N_2(a, \mu(b)) = 1 + \|[a, \mu(b)]\|_{\mathcal{B}} + N_2(a, \mu(b)), \quad (8.13)$$

which concludes the proof of (8.11) by induction, since  $\theta_a(b) = \theta_a(\mu(b)) + 1$ .  $\square$

**Proposition 8.15.** *Let  $a < b \in \mathcal{B}$  such that  $(a, b) \notin \mathcal{B}$  and  $\lambda^2(b) \leq a$ . Then*

$$\|[a, b]\|_{\mathcal{B}} \leq (\theta - 2)2^{\theta-2} + \theta \quad \text{and} \quad N_2(a, b) \leq (\theta - 2)2^{\theta-2} + 1, \quad \text{where } \theta = \theta_a(b). \quad (8.14)$$

*Proof.* We proceed by induction on  $\theta_a(b) \geq 2$ . Initialization is proved by Lemma 8.13. We perform the inductive step. Assume  $\theta_a(b) \geq 3$ . Then  $\theta_a(\mu(b)) = \theta_a(b) - 1 \geq 2$  thus  $a < \lambda(\mu(b))$ ,  $\lambda(\mu(b)) \notin X$  and  $\lambda^2(\mu(b))$  is well defined. We study both terms of Jacobi's identity (8.12).

*First term.* We have  $(a, \lambda(b)) < \mu(b)$  by Definition 8.2 because  $a < \lambda(\mu(b))$ .

- If  $\lambda(\mu(b)) \leq (a, \lambda(b))$  then  $c := ((a, \lambda(b)), \mu(b)) \in \mathcal{B}$  and  $\theta_{\lambda(b)}(c) \in \{1, 2\}$ .
- If  $(a, \lambda(b)) < \lambda(\mu(b))$ , we apply Proposition 8.14 to the bracket  $[(a, \lambda(b)), \mu(b)]$ . Indeed  $\lambda^2(\mu(b)) \leq \lambda^2(b) \leq a < (a, \lambda(b))$ . By Lemma 4.8,  $\theta_{(a, \lambda(b))}(\mu(b)) \leq \theta_a(\mu(b)) = \theta_a(b) - 1$ . Therefore Proposition 8.14 gives  $\|[a, \lambda(b)], \mu(b)\|_{\mathcal{B}} \leq 2^{\theta_a(b)-2}$  and  $N_2((a, \lambda(b)), \mu(b)) \leq 2^{\theta_a(b)-2} - 1$ . Any  $c \in \text{supp}[(a, \lambda(b)), \mu(b)]$  satisfies  $\theta_{\lambda(b)}(c) \leq \theta_{\lambda(\mu(b))}(c)$  by Lemma 4.8.

*Second term.* By Theorem 8.9,  $[a, \mu(b)] = \sum \alpha_d E(d)$  where the sum ranges over  $d \in \mathcal{B}$  such that  $a < \lambda(d) < \mu(b)$ ,  $\lambda(\mu(b)) < d$  and  $\theta_{\lambda(\mu(b))}(d) \leq 2$  using Lemma 8.12. Since  $\lambda^2(\mu(b)) \leq \lambda^2(b) \leq a$ , the induction hypothesis applies to  $[a, \mu(b)]$ . Let  $d \in \text{supp}[a, \mu(b)]$ . By Lemma 4.8,  $\theta_{\lambda(b)}(d) \leq \theta_{\lambda(\mu(b))}(d)$ . Moreover, since  $\lambda^2(b) \leq a < \lambda(d)$  (by (2.2)),  $\lambda(b) < d$  by Definition 8.2.

- If  $\theta_{\lambda(b)}(d) = 1$ ,  $c := (\lambda(b), d) \in \mathcal{B}$  and  $\theta_{\lambda(b)}(c) = 1$ .
- If  $\theta_{\lambda(b)}(d) = 2$ , then  $\|[\lambda(b), d]\|_{\mathcal{B}} \leq 2$ .

*Conclusion.* This analysis proves  $N_2(a, b) \leq 2^{\theta_a(b)-2} - 1 + 2N_2(a, \mu(b))$  and that

$$\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta_a(b)-2} + N_1(a, \mu(b)) + 2N_2(a, \mu(b)) = 2^{\theta_a(b)-2} + \|[a, \mu(b)]\|_{\mathcal{B}} + N_2(a, \mu(b)), \quad (8.15)$$

which concludes the proof of (8.14) by induction, since  $\theta_a(b) = \theta_a(\mu(b)) + 1$ .  $\square$

## 8.6 General estimate

For  $a < b \in \mathcal{B}$ , we decompose  $\theta_a(b) = n_a(b) + \nu_a(b)$  where  $n_a(b)$  (resp.  $\nu_a(b)$ ) denotes the number of leaves of  $T_a(b)$  different from  $X_1$  (resp. equal to  $X_1$ ). We start by stating elementary monotony properties of these quantities. We then prove a bound on  $\|[a, b]\|_{\mathcal{B}}$  involving these quantities in Theorem 8.18, which yields Corollary 8.19, a slightly weaker version of our main goal, Theorem 1.7.

**Lemma 8.16.** *Let  $a \leq h < b \in \mathcal{B}$ . Then  $n_h(b) \leq n_a(b)$  and  $\nu_h(b) \leq \nu_a(b)$ .*

*Proof.* We proceed by induction on  $\theta_a(b)$ .

*Initialization for  $\theta_a(b) = 1$ .* Then either  $b = X_1$  or  $\lambda(b) \leq a \leq h$ . In both cases,  $T_a(b) = b = T_h(b)$  so  $(n_a(b), \nu_a(b)) = (n_h(b), \nu_h(b))$ .

*Induction for  $\theta_a(b) \geq 2$ .* Then  $b \neq X_1$  and  $n_a(b) \geq 1$ .

- *Case  $\lambda(b) \leq h$ .* Then  $n_h(b) = 1 \leq n_a(b)$  and  $\nu_h(b) = 0 \leq \nu_a(b)$ .
- *Case  $h < \lambda(b)$ .* By the induction hypothesis,  $n_h(\lambda(b)) \leq n_a(\lambda(b))$  and  $n_h(\mu(b)) \leq n_a(\mu(b))$ . Therefore,  $n_h(b) \leq n_a(b)$ . The same argument shows that  $\nu_h(b) \leq \nu_a(b)$ .  $\square$

Under additional assumptions, we now prove that their weighted sum is strictly decreasing. This remark plays a key role in the proof of Theorem 8.18 and even more in its refined version for particular brackets stated in Proposition 8.23.

**Lemma 8.17.** *Let  $a < h < b \in \mathcal{B}$ . Assume that either  $h = (a, X_1)$  or  $a < \lambda(h)$ . Then*

$$2n_h(b) + \nu_h(b) \leq n_a(b) + \nu_a(b). \quad (8.16)$$

*In particular, if  $b \neq X_1$ ,  $2n_h(b) + \nu_h(b) < 2n_a(b) + \nu_a(b)$ .*

*Proof.* We proceed by induction on  $\theta_a(b)$ .

*Initialization for  $\theta_a(b) = 1$ .* Then either  $b = X_1$  or  $\lambda(b) \leq a$ . If  $b = X_1$ ,  $n_a(b) = n_h(b) = 0$  and  $\nu_a(b) = \nu_h(b) = 1$ , so (8.16) holds. Let us prove that the case  $b \neq X_1$  with  $\lambda(b) \leq a$  cannot happen. We deduce from  $h < b$  that  $\lambda(h) \leq \lambda(b)$ . In the case  $h = (a, X_1)$ , this relation leads to  $\lambda(b) = a$  and  $b = (a, \mu(b)) \leq (a, X_1) = h$ , which is a contradiction. In the case  $a < \lambda(h)$ , this relation leads to  $a < \lambda(h) \leq \lambda(b) \leq a$ , also a contradiction.

*Induction for  $\theta_a(b) \geq 2$ .* Then  $n_a(b) \geq 1$ .

- *Case  $\lambda(b) \leq h$ .* Then  $n_h(b) = 1$ ,  $\nu_h(b) = 0$  thus (8.16) holds because the right-hand side is equal to  $\theta_a(b) \geq 2$ .
- *Case  $h < \lambda(b)$ .* By the induction hypothesis  $2n_h(\lambda(b)) + \nu_h(\lambda(b)) \leq n_a(\lambda(b)) + \nu_a(\lambda(b))$  and  $2n_h(\mu(b)) + \nu_h(\mu(b)) \leq n_a(\mu(b)) + \nu_a(\mu(b))$  and (8.16) follows by summing these two inequalities.  $\square$

**Theorem 8.18.** *For every  $a < b \in \mathcal{B}$ ,*

$$\|[a, b]\|_{\mathcal{B}} \leq F_{2n_a(b) + \nu_a(b)}. \quad (8.17)$$

*Proof.* We proceed, as in Proposition 4.17, by induction on  $\theta_a(b)$ , by applying the classical rewriting scheme. We refer to the proof of Proposition 4.17 for a detailed justification of why the induction on  $\theta_a(b)$  is legitimate in this setting.

*Case 1:  $\theta_a(b) = 1$  i.e. either  $b = X_1$  or  $\lambda(b) \leq a$ .* Then  $\|[a, b]\|_{\mathcal{B}} = 1$ , which is the value in the right-hand side of (8.17) both when  $b = X_1$  and when  $n_a(b) = 1$ ,  $\nu_a(b) = 0$  since  $F_1 = F_2 = 1$ .

From now on, we assume  $\theta_a(b) > 1$ , so  $b \notin X$  and  $a < \lambda(b)$ . Thus  $\lambda(b) \notin X$ .

*Case 2:  $\theta_a(\lambda(b)) = 1$  i.e.  $\lambda^2(b) \leq a < \lambda(b)$ .* We apply Proposition 8.15. Combining (8.14) and (D.4) yields  $\|[a, b]\|_{\mathcal{B}} \leq F_{2\theta_a(b) - 1}$ . Moreover  $2\theta_a(b) - 1 = 2n_a(b) + \nu_a(b) + (\nu_a(b) - 1)$ . By Lemma 8.10,  $\nu_a(b) \in \{0, 1\}$ , so  $F_{2\theta_a(b) - 1} \leq F_{2n_a(b) + \nu_a(b)}$ , and (8.17) holds.

From now on, we also assume  $a < \lambda^2(b)$ . Hence  $\theta_a(\lambda(b)) > 1$ . The proof relies on the Jacobi decomposition (8.4). We introduce the notations  $n_1 := n_a(\lambda(b))$ ,  $\nu_1 := \nu_a(\lambda(b))$ ,  $n_2 := n_a(\mu(b))$ ,  $\nu_2 := \nu_a(\mu(b))$ . In particular  $n_1 + \nu_1 > 1$ . By the induction hypothesis,

$$\|[a, \lambda(b)]\|_{\mathcal{B}} \leq F_{2n_1 + \nu_1}, \quad (8.18)$$

$$\|[a, \mu(b)]\|_{\mathcal{B}} \leq F_{2n_2 + \nu_2}. \quad (8.19)$$

*Case 3:  $\theta_a(\lambda(b)) \geq 2$  i.e.  $a < \lambda^2(b)$  and  $\mu(b) = X_1$ .* Concerning the first term of (8.4), for each  $d \in \text{supp}[a, \lambda(b)]$ ,  $(d, \mu(b)) = (d, X_1) \in \mathcal{B}$ , so (8.18) yields  $\|[a, \lambda(b)], \mu(b)\|_{\mathcal{B}} \leq F_{2n_1 + \nu_1}$ . Concerning the second term of (8.4),  $d := (a, \mu(b)) = (a, X_1) \in \mathcal{B}$ . If  $\lambda(b) < d$ , then  $(\lambda(b), d) \in \mathcal{B}$  because  $a < \lambda(b)$ . If  $d < \lambda(b)$ , the induction hypothesis yields  $\|[a, \mu(b)], \lambda(b)\|_{\mathcal{B}} \leq F_{2n'_1 + \nu'_1}$  where  $n'_1 := n_d(\lambda(b))$  and  $\nu'_1 := \nu_d(\lambda(b))$ . By Lemma 8.17 (applied to  $a \leftarrow a$ ,  $h \leftarrow d = (a, X_1)$ ,  $b \leftarrow \lambda(b)$ ),  $2n'_1 + \nu'_1 < 2n_1 + \nu_1$ . Hence  $\|[a, \mu(b)], \lambda(b)\|_{\mathcal{B}} \leq F_{2n_1 + \nu_1 - 1}$ . Summing our estimates for both terms of (8.4) provides

$$\|[a, b]\|_{\mathcal{B}} \leq F_{2n_1 + \nu_1} + F_{2n_1 + \nu_1 - 1} = F_{2n_1 + (\nu_1 + 1)}, \quad (8.20)$$

which proves (8.17) when  $\mu(b) = X_1$  since  $n_a(b) = n_1$  and  $\nu_a(b) = \nu_1 + 1$ .

*Case 4:  $\theta_a(\lambda(b)) \geq 2$  i.e.  $a < \lambda^2(b)$ ,  $\mu(b) \notin X$  and  $\theta_a(\mu(b)) = 1$ .* Since  $\mu(b) \neq X_1$ ,  $n_2 = 1$  and  $\nu_2 = 0$ . We consider both terms of (8.4).

- For the first term, let  $d \in \text{supp}[a, \lambda(b)]$ . Since  $(a, \lambda(b)) \notin \mathcal{B}$ , hence, by (2.2),  $a < \lambda(d)$ . Since  $\theta_a(\mu(b)) = 1$ ,  $\lambda(\mu(b)) \leq a < \lambda(d)$  so  $\mu(b) < d$  automatically. Then  $(\mu(b), d) \in \mathcal{B}$  since, by Theorem 8.9,  $\lambda(d) < \lambda(b) < \mu(b)$ . So  $\|[[a, \lambda(b)], \mu(b)]\|_{\mathcal{B}} \leq F_{2n_1+\nu_1}$ .
- For the second term, when  $\theta_a(\mu(b)) = 1$ ,  $d := (a, \mu(b)) \in \mathcal{B}$ .  
If  $\lambda(b) < d$ ,  $(\lambda(b), d) \in \mathcal{B}$  since  $a < \lambda(b)$ , so  $\|[\lambda(b), [a, \mu(b)]]\|_{\mathcal{B}} = 1$ .  
If  $d < \lambda(b)$ , by the induction hypothesis,  $\|[d, \lambda(b)]\|_{\mathcal{B}} \leq F_{2n'_1+\nu'_1}$  where  $n'_1 := n_d(\lambda(b))$  and  $\nu'_1 := \nu_d(\lambda(b))$ . By Lemma 8.16 (applied to  $a \leftarrow a$ ,  $h \leftarrow d$ ,  $b \leftarrow \lambda(b)$ ),  $n'_1 \leq n_1$  and  $\nu'_1 \leq \nu_1$ . So  $\|[\lambda(b), [a, \mu(b)]]\|_{\mathcal{B}} \leq F_{2n_1+\nu_1}$ .

As a conclusion, using (D.1),

$$\|[a, b]\|_{\mathcal{B}} \leq 2F_{2n_1+\nu_1} \leq F_{2n_1+\nu_1+2} = F_{2(n_1+1)+\nu_1}, \quad (8.21)$$

which proves (8.17) in this case.

*Case 5:*  $\theta_a(\lambda(b)) \geq 2$  i.e.  $a < \lambda^2(b)$ ,  $\mu(b) \notin X$  and  $\theta_a(\mu(b)) \geq 2$ .

- We consider the first term of (8.4). Let  $d \in \text{supp}[a, \lambda(b)]$ . Since  $(a, \lambda(b)) \notin \mathcal{B}$ ,  $a < \lambda(d)$ , by (2.2).  
If  $\mu(b) < d$ , then  $\|[\mu(b), d]\|_{\mathcal{B}} = 1$  since, by Theorem 8.9,  $\lambda(d) < \lambda(b) < \mu(b)$ .  
If  $d < \mu(b)$ , by the induction hypothesis,  $\|[d, \mu(b)]\|_{\mathcal{B}} \leq F_{2n'_2+\nu'_2}$  where  $n'_2 := n_d(\mu(b))$  and  $\nu'_2 := \nu_d(\mu(b))$ . Since  $a < \lambda(d)$ , we can apply Lemma 8.17 to  $(a \leftarrow a, h \leftarrow d, b \leftarrow \mu(b) \neq X_1)$ , which proves that  $2n'_2 + \nu'_2 \leq 2n_2 + \nu_2 - n_2$ .  
Hence, summing over  $d$  and using (8.18), this provides the estimate

$$\|[[a, \lambda(b)], \mu(b)]\|_{\mathcal{B}} \leq F_{2n_1+\nu_1} F_{2n_2+\nu_2-n_2}. \quad (8.22)$$

- We consider the second term of (8.4). Let  $d \in \text{supp}[a, \mu(b)]$ . Since we already covered the case  $\theta_a(\mu(b)) = 1$ ,  $(a, \mu(b)) \notin \mathcal{B}$ , hence, by (2.2),  $a < \lambda(d)$ .  
If  $\lambda(b) < d$ , we know from the proof of Theorem 8.9 that  $\|[\lambda(b), d]\|_{\mathcal{B}} \leq 2$ .  
If  $d < \lambda(b)$ , by the induction hypothesis,  $\|[d, \lambda(b)]\|_{\mathcal{B}} \leq F_{2n'_1+\nu'_1}$  where  $n'_1 = n_d(\lambda(b))$  and  $\nu'_1 = \nu_d(\lambda(b))$ . Since  $a < \lambda(d)$ , we can apply Lemma 8.17 to  $(a \leftarrow a, h \leftarrow d, b \leftarrow \mu(b))$ , which proves that  $2n'_1 + \nu'_1 \leq 2n_1 + \nu_1 - 1$ .

Hence, summing over  $d$  and using (8.19), this provides the estimate

$$\|[\lambda(b), [a, \mu(b)]]\|_{\mathcal{B}} \leq F_{2n_2+\nu_2} \max\{2, F_{2n_1+\nu_1-n_1}\}. \quad (8.23)$$

*Subcase 5.1:*  $n_1 = 1$  and  $\nu_1 = 1$ . Then the maximum in the right-hand side of (8.23) is 2 because  $F_2 = 1$ . Summing (8.22) and (8.23), and using (D.1) yields

$$\|[a, b]\|_{\mathcal{B}} \leq F_3 F_{2n_2+\nu_2-1} + 2F_{2n_2+\nu_2} = 2F_{2n_2+\nu_2+1} \leq F_{2n_2+\nu_2+3}, \quad (8.24)$$

which proves (8.17) in this case since  $2n_1 + \nu_1 = 3$ .

*Subcase 5.2:*  $n_1 = 2$  and  $\nu_1 = 0$ . Then the maximum in the right-hand side of (8.23) is 2 because  $F_2 = 1$ . Summing (8.22) and (8.23), and using (D.2) yields

$$\|[a, b]\|_{\mathcal{B}} \leq F_4 F_{2n_2+\nu_2-1} + 2F_{2n_2+\nu_2} \leq F_{2n_2+\nu_2+3} = F_{2(n_1+n_2)+(\nu_1+\nu_2)-1}, \quad (8.25)$$

which proves (8.17) in this case.

*Subcase 5.3:*  $n_1 + \nu_1 > 2$ . Then the maximum in the right-hand side of (8.23) is  $F_{n_1+\nu_1}$  (with equality when  $n_1 + \nu_1 = 3$ ). Summing (8.22) and (8.23), and using (D.2) yields,

$$\|[a, b]\|_{\mathcal{B}} \leq F_{2n_1+\nu_1} F_{2n_2+\nu_2-1} + F_{2n_1+\nu_1-1} F_{2n_2+\nu_2} \leq F_{2(n_1+n_2)+(\nu_1+\nu_2)-1}, \quad (8.26)$$

which proves (8.17) in this case.  $\square$



An immediate consequence is that Theorem 8.18 allows to prove a slightly weaker version of Theorem 1.7 (with an estimate by  $F_{n-1}$  instead of  $F_{n-2}$ ).

**Corollary 8.19.** *Let  $a < b \in \mathcal{B}$  with  $n := |a| + |b| \geq 3$ . Then  $\|[a, b]\|_{\mathcal{B}} \leq F_{n-1}$ .*

*Proof.* Let  $c \in L_{\mathcal{B}}(T_a(b))$ . So  $c \neq X_0$ . If  $c \neq X_1$ ,  $|c| \geq 2$ . Thus,  $|b| \geq 2n_a(b) + \nu_a(b)$ . Since  $|a| \geq 1$ ,  $n - 1 = |a| + |b| - 1 \geq 2n_a(b) + \nu_a(b)$  so (8.17) proves that  $\|[a, b]\|_{\mathcal{B}} \leq F_{n-1}$ .  $\square$

**Remark 8.20.** *As soon as  $|a| > 1$  or  $|b| > 2n_a(b) + \nu_a(b)$ , estimate (8.17) even proves that  $\|[a, b]\|_{\mathcal{B}} \leq F_{n-2}$ . Hence, to conclude the proof of Theorem 1.7, it only remains to prove that, when  $|a| = 1$  (so  $a = X_0$ ) and  $|b| = 2n_a(b) + \nu_a(b)$ , one has  $\|[a, b]\|_{\mathcal{B}} \leq F_{|b|-1}$ . This is the goal of Section 8.7 (see Proposition 8.23).*

## 8.7 Enhanced estimate for brackets with simple leaves

In this paragraph, we focus on the special case when  $a = X_0$  and  $b \in \mathcal{B} \setminus X$  has simple leaves in the sense that  $L_{\mathcal{B}}(T_a(b)) \subset A$  where  $A := \{(X_0, X_1), X_1\}$ . Such brackets are those satisfying  $|b| = 2n_a(b) + \nu_a(b)$  and  $|b| \geq 2$ . In this setting, we prove in Proposition 8.23 that the index of the Fibonacci bound (8.17) can be decreased.

We start with a short lemma on structural properties of such brackets, then prove the desired estimate in a particular case before proceeding to the main result.

**Lemma 8.21.** *Let  $b \in \mathcal{B} \cap \mathfrak{I}(\text{Br}(A))$ . Then  $b \geq (X_0, X_1)$  with equality iff ( $b \neq X_1$  and  $\lambda(b) = X_0$ ).*

*Proof.* This is an immediate consequence of Lemma 1.41 and the minimality of  $X_0$  in  $\mathcal{B}$ .  $\square$

**Lemma 8.22.** *Let  $a := X_0$  and  $b \in \mathcal{B} \setminus X$  such that  $|b| = 2n_a(b) + \nu_a(b) \geq 2$ . Assume that  $\lambda^2(b) = X_0 < \lambda(b)$ . Then  $b = \text{ad}_{(X_0, X_1)}^{n_a(b)}(X_1)$  and  $\|[a, b]\|_{\mathcal{B}} \leq F_{2n_a(b) + \nu_a(b) - 1}$ .*

*Proof.* Let us prove both claims successively.

*First item.* Since  $\theta_a(b) > 1$ ,  $\lambda(b) = \mathfrak{I}(\lambda(T_a(b)))$  so  $\lambda(b) \in \mathfrak{I}(\text{Br}(A))$ . By Lemma 8.21,  $\lambda(b) = (X_0, X_1)$ . Let  $n \geq 1$  such that  $b = \text{ad}_{(X_0, X_1)}^n(c)$  with either  $c = X_1$  or  $\lambda(c) < (X_0, X_1)$ . Let us check that  $c \neq X_1$  cannot occur. If  $c \neq X_1$  then  $c \in \mathfrak{I}(\text{Br}(A))$ . If  $c = (X_0, X_1)$ ,  $b \notin \mathcal{B}$  since  $\mathfrak{E}(b) = 0$ . Otherwise  $\lambda(c) \in \mathfrak{I}(\text{Br}(A))$  so  $\lambda(c) \geq (X_0, X_1)$  by Lemma 8.21 which contradicts  $\lambda(c) < (X_0, X_1)$ . So  $c = X_1$  and  $n = n_a(b)$ .

*Size estimate.* Since  $\lambda^2(b) = X_0 = a$ , estimate (8.14) of Proposition 8.15 yields  $\|[a, b]\|_{\mathcal{B}} \leq (n-1)2^{n-1} + n + 1$  where  $n = n_a(b)$ . Moreover,  $|b| = 2n_a(b) + \nu_a(b) = 2n + 1$  so  $F_{|b|-1} = F_{2n}$ . This proves the claimed estimate for  $n \geq 8$  thanks to (D.6). For  $n \in \llbracket 1, 7 \rrbracket$ , the estimate  $\|[a, b]\|_{\mathcal{B}} \leq F_{2n}$  can be checked using computer algebra for these short explicit brackets. It could alternatively be proved from the enhanced bound of Corollary 8.32 and (D.5).  $\square$

**Proposition 8.23.** *Let  $a := X_0$  and  $b \in \mathcal{B} \setminus X$  such that  $|b| = 2n_a(b) + \nu_a(b) \geq 2$ . Then*

$$\|[a, b]\|_{\mathcal{B}} \leq F_{2n_a(b) + \nu_a(b) - 1}. \quad (8.27)$$

*Proof.* The proof follows along the same lines as the proof of Theorem 8.18, taking advantage of the additional assumptions.

*Case 1:*  $\theta_a(b) = 1$  i.e. either  $b = X_1$  or  $\lambda(b) \leq a$ . Then  $\|[a, b]\|_{\mathcal{B}} = 1$  and  $F_{2n_a(b) + \nu_a(b) - 1} \geq F_1 = 1$  since we assumed that  $|b| = 2n_a(b) + \nu_a(b) \geq 2$ . So (8.27) holds in this case.

*Case 2:*  $\theta_a(\lambda(b)) = 1$  i.e.  $\lambda^2(b) \leq a < \lambda(b)$ . Since  $\lambda^2(b) \leq a = X_0$  and  $X_0$  is minimal in  $\mathcal{B}$ ,  $\lambda^2(b) = X_0$ . Then (8.27) follows from Lemma 8.22.

From now on, we assume  $a < \lambda^2(b)$  (so  $\theta_a(\lambda(b)) > 1$ ). We reuse the notations  $n_1 := n_a(\lambda(b))$ ,  $\nu_1 := \nu_a(\lambda(b))$ ,  $n_2 := n_a(\mu(b))$ ,  $\nu_2 := \nu_a(\mu(b))$ .

*Case 3:*  $\theta_a(\lambda(b)) \geq 2$  i.e.  $a < \lambda^2(b)$  and  $\mu(b) = X_1$ . Then  $n_a(b) = n_1$  and  $\nu_a(b) = \nu_1 + 1$ .

- If  $n_a(b) = 1$ , then  $b = \text{ad}_{X_1}^{\nu_a(b)}(X_0, X_1) = \text{ad}_{X_1}^{\nu_a(b)+1}(X_0)$ . Since  $r(X_0, X_1) = +\infty$ , (5.10) yields  $\| [a, b] \|_{\mathcal{B}} = F_{\nu_a(b)+1}$ . Since  $n_a(b) = 1$ ,  $\nu_a(b) + 1 = 2n_a(b) + \nu_a(b) - 1$  so (8.27) is proved.
- If  $n_a(b) = n_1 > 1$ , we write (8.4). Concerning the first term of (8.4), we apply (8.27) to  $a \leftarrow a$  and  $b \leftarrow \lambda(b)$  (the latter indeed satisfies the assumptions of the proposition). Hence  $\| [a, \lambda(b)] \|_{\mathcal{B}} \leq F_{2n_1+\nu_1-1}$ . For each  $d \in \text{supp}[a, \lambda(b)]$ ,  $(d, \mu(b)) = (d, X_1) \in \mathcal{B}$ , so  $\| [[a, \lambda(b)], \mu(b)] \|_{\mathcal{B}} \leq F_{2n_1+\nu_1-1}$ . Concerning the second term of (8.4),  $d := (a, \mu(b)) = (a, X_1) \in \mathcal{B}$ . If  $\lambda(b) < d$ , then  $(\lambda(b), d) \in \mathcal{B}$  because  $a < \lambda(b)$ . If  $d < \lambda(b)$ , (8.17) yields  $\| [[a, \mu(b)], \lambda(b)] \|_{\mathcal{B}} \leq F_{2n'_1+\nu'_1}$  where  $n'_1 := n_d(\lambda(b))$  and  $\nu'_1 := \nu_d(\lambda(b))$ . By Lemma 8.17 (applied to  $a \leftarrow a$ ,  $h \leftarrow d = (a, X_1)$ ,  $b \leftarrow \lambda(b)$ ),  $2n'_1+\nu'_1 \leq 2n_1+\nu_1-n_1 \leq 2n_1+\nu_1-2$ . Hence  $\| [[a, \mu(b)], \lambda(b)] \|_{\mathcal{B}} \leq F_{2n_1+\nu_1-2}$ . Summing our estimates for both terms of (8.4) provides  $\| [a, b] \|_{\mathcal{B}} \leq F_{2n_1+\nu_1-1} + F_{2n_1+\nu_1-2} = F_{2n_1+\nu_1}$ , which proves (8.27) when  $\mu(b) = X_1$  since  $n_a(b) = n_1$  and  $\nu_a(b) = \nu_1 + 1$ .

*Case 4:*  $\theta_a(\lambda(b)) \geq 2$  i.e.  $a < \lambda^2(b)$ ,  $\mu(b) \notin X$  and  $\theta_a(\mu(b)) = 1$ . This case does not happen. Indeed, since  $\mu(b) \neq X_1$ ,  $\theta_a(\mu(b)) = 1$  implies that  $\lambda(\mu(b)) \leq a = X_0$  so, by Lemma 8.21,  $\mu(b) = (X_0, X_1)$ . But, by Lemma 8.21,  $\lambda(b) \geq (X_0, X_1)$ , which contradicts  $\lambda(b) < \mu(b)$ .

*Case 5:*  $\theta_a(\lambda(b)) \geq 2$  i.e.  $a < \lambda^2(b)$ ,  $\mu(b) \notin X$  and  $\theta_a(\mu(b)) \geq 2$ . We proceed exactly as in the *general case* of the proof of Theorem 8.18. When  $(n_1, \nu_1) = (2, 0)$  or  $n_1 + \nu_1 > 2$ , estimates (8.25) and (8.26) directly prove the conclusion (8.27). When  $n_1 = 1$  and  $\nu_1 = 1$ ,  $\lambda(b) = ((X_0, X_1), X_1)$  so  $[a, \lambda(b)] = \text{E}(d)$  where  $d := (\text{ad}_{X_0}^2(X_1), X_1) \in \mathcal{B}$  and (8.22) becomes

$$\| [[a, \lambda(b)], \mu(b)] \|_{\mathcal{B}} \leq F_{2n_2+\nu_2-n_2}. \quad (8.28)$$

Together with (8.23), this yields

$$\| [a, b] \|_{\mathcal{B}} \leq F_{2n_2+\nu_2-1} + 2F_{2n_2+\nu_2} = F_{2n_2+\nu_2+2} = F_{2(n_1+n_2)+(\nu_1+\nu_2)-1}, \quad (8.29)$$

which concludes the proof of (8.27).  $\square$

As mentioned in Remark 8.20, the bound (8.27) of Proposition 8.23 concludes the proof of Theorem 1.7 with the optimal index of the Fibonacci sequence.

## 8.8 Optimal $\theta$ -based estimate

In this paragraph, we derive an optimal bound for  $\| [a, b] \|_{\mathcal{B}}$  with respect to  $\theta_a(b)$ . We define the following integer-valued function which will correspond to the optimal  $\theta$ -based estimate of the structure constants of  $\mathcal{B}$ :

$$A(\theta) := (\theta - 3)2^{\theta-2} + \theta + 1. \quad (8.30)$$

First, we give in Section 8.8.1 examples of brackets  $a < b \in \mathcal{B}$  such that  $\| [a, b] \|_{\mathcal{B}} = A(\theta_a(b))$ , illustrating that the optimal  $\theta$ -based estimate within  $\mathcal{B}$  is in particular larger than  $2^{\theta-1}$ . We discuss in Section 8.8.2 on the apparent paradox that  $\mathcal{B}$  is minimal for length-based estimates but not for  $\theta$ -based ones due to the previous examples.

Since this question interested us, was involved in an earlier proof of Theorem 8.18 and is of independent interest for a deeper understanding of  $\mathcal{B}$ , we then prove that  $\| [a, b] \|_{\mathcal{B}} \leq A(\theta_a(b))$  for every  $a < b \in \mathcal{B}$ . We start in Section 8.8.3 by the particular case where  $T_a(b)$  is right-nested (in the sense of Lemma 8.10). Eventually, we extend this bound to all brackets in Section 8.8.4.

### 8.8.1 An example of brackets attaining the upper bound

**Proposition 8.24.** *Let  $p \geq 1$ ,  $m \geq 2$ ,  $a := (X_0, X_1)$ ,  $h := \text{ad}_a^m(X_1)$  and  $b := \text{ad}_h^p(X_1)$ . Then  $\theta_a(b) = p + 1$  and  $\| [a, b] \|_{\mathcal{B}} = A(\theta_a(b))$ .*

*Proof.* We have

$$\lambda(h) = a < h \quad \text{and} \quad (a, h) < h. \quad (8.31)$$

In particular,  $(a, b) \notin \mathcal{B}$ ,  $(a, h) \in \mathcal{B}$ ,  $(a, X_1) \in \mathcal{B}$  thus  $\theta_a(b) = p + 1$  and our goal is to prove  $\|[a, b]\|_{\mathcal{B}} = (p - 2)2^{p-1} + p + 2$ . Using (5.4) with  $r = 2$  and (5.5) (although these formulas are stated for a bracket of the form  $[X_0, \text{ad}_{X_1}^r(X_2)]$ , their derivation solely relies on Jacobi's identity, so remains just as valid for our bracket  $[a, \text{ad}_h^p(X_1)]$ ), we get

$$[a, b] = \text{ad}_h^p([a, X_1]) + p \text{ad}_h^{p-1}([(a, h), X_1]) + \sum_{s=2}^p (s-1) \binom{p}{s} [\text{ad}_h^{s-2}((a, h), h), \text{ad}_h^{p-s}(X_1)]. \quad (8.32)$$

Let us prove that the  $(p+1)$  brackets in the right-hand side are (evaluations of) different elements of  $\mathcal{B}$ . The first two ones are by (8.31). Let  $s \in \llbracket 2, p \rrbracket$ . We have  $\text{ad}_h^{s-2}((a, h), h) < \text{ad}_h^{p-s}(X_1)$ , either because  $X_1$  is maximal when  $(p-s) \leq (s-2)$ , or because  $(a, h) < h$  when  $(s-2) < (p-s)$ . Thus  $(\text{ad}_h^{s-2}((a, h), h), \text{ad}_h^{p-s}(X_1)) \in \mathcal{B}$ , because, when  $s \leq p-1$ , we have  $h < \text{ad}_h^{s-2}((a, h), h)$  (this relation holds because  $\lambda(h) = a < (a, h)$  when  $s = 2$ ). In conclusion,

$$\|[a, b]\|_{\mathcal{B}} = 1 + p + \sum_{s=2}^p (s-1) \binom{p}{s} = 1 + p + \sum_{s=2}^p p \binom{p-1}{s-1} = (2^p - 1 - p), \quad (8.33)$$

which concludes the proof. (One could wish to rely on (5.4) and on Proposition 3.5 with the triple  $\{a, h, X_1\}$  to avoid checking that the brackets are indeed part of the basis, but, unfortunately, although this set is alphabetic, it is not free since  $h = \text{ad}_a^m(X_1)$ ).  $\square$

### 8.8.2 Discussion on the minimality of our basis

Proposition 8.24 proves that, for every  $\theta \geq 3$ , there exist  $a < b \in \mathcal{B}$  with  $\theta_a(b) = \theta$  and  $\|[a, b]\|_{\mathcal{B}} = A(\theta_a(b)) > 2^{\theta_a(b)-1}$  (and, asymptotically,  $A(\theta) \gg 2^{\theta-1}$ ). Hence, although the structure constants of the basis constructed in Section 8.2 have a minimal growth with respect to the length of the involved brackets, they do not have minimal growth with respect to  $\theta_a(b)$ . Indeed, for length-compatible Hall sets (see Theorem 6.7) and for the Lyndon basis (see Theorem 7.5), one has  $\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta_a(b)-1}$ , which is the minimal  $\theta$ -based bound, due to the lower-bound examples of Proposition 5.7.

In Proposition 4.7, we proved that for any Hall set with  $|X| = 2$ ,  $\theta_a(b) \leq |b| - 1$  and that this bound was attained within each Hall set. Hence, the apparent paradox between the fact that  $\mathcal{B}$  is somehow length-minimal but not  $\theta$ -minimal does not come from a better estimation of  $\theta_a(b)$  from  $|b|$ . However, the brackets such that  $\theta_a(b) = |b| - 1$  are  $a = X_0$  and  $b = \text{ad}_{X_1}^n(X_0)$  for some  $n \in \mathbb{N}^*$ , which does not match the form of the brackets attaining equality in Proposition 8.24, which explains why there is no contradiction.

### 8.8.3 Sharp bound for right-nested brackets

The goal of this paragraph is to prove Theorem 8.31, which improves the bound given in Proposition 8.15 from  $(\theta - 2)2^{\theta-2} + \theta$  down to  $A(\theta)$  (already for  $\theta = 3$ , the latter is strictly smaller than the former). The proof builds upon the method developed in Section 8.5. The tighter bound comes notably from the identification of particular elements of the support that “change type” (from being part of  $N_2$  to being part of  $N_1$ ) at each induction step, thereby avoiding a two-fold increase of the associated part of the norm.

We start by a slight precision in the conclusions of Lemma 8.13.

**Lemma 8.25.** *Let  $a < b \in \mathcal{B}$  with  $\theta_a(b) = 2$ . Then  $\|[a, b]\|_{\mathcal{B}} \leq 2$  and  $N_2(a, b) \leq 1$ . Moreover, when  $N_2(a, b) = 1$ , the unique  $c \in \text{supp}[a, b]$  such that  $\theta_{\lambda(b)}(c) = 2$  satisfies  $\lambda(c) \leq (a, \lambda(b))$ .*

*Proof.* The first assertions are proved in Lemma 8.13. When  $N_2(a, b) = 1$ , with the notations of the proof of Lemma 8.13, the unique  $c \in \text{supp}[a, b]$  such that  $\theta_{\lambda(b)}(c) = 2$  is named  $c_1$  and satisfies  $\lambda(c_1) = (a, \lambda(b))$  or  $\lambda(c_1) = \mu(b)$  (when  $\mu(b) < (a, \lambda(b))$ ). In both cases,  $\lambda(c) \leq (a, \lambda(b))$ .  $\square$

We now wish to weaken the assumption  $\lambda^2(b) < a$  of Proposition 8.14 down to  $\lambda(b) < (a, \lambda(b))$ . We start by checking that this assumption is stable in the induction process going from  $b$  to  $\mu(b)$ .

**Lemma 8.26.** *Let  $a < b \in \mathcal{B}$  such that  $\theta_a(b) \geq 3$ ,  $(a, \lambda(b)) \in \mathcal{B}$  and  $\lambda(b) < (a, \lambda(b))$ . Then  $(a, \lambda(\mu(b))) \in \mathcal{B}$  and  $\lambda(\mu(b)) < (a, \lambda(\mu(b)))$ .*

*Proof.* The assumption  $\theta_a(b) \geq 3$  implies  $(a, b) \notin \mathcal{B}$  thus  $b \neq X_1$ , so  $\lambda(b)$  makes sense and  $a < \lambda(b)$ . Since  $(a, \lambda(b)) \in \mathcal{B}$  then  $\lambda(b) \neq X_0$ , so  $\lambda^2(b)$  is well defined and  $\lambda^2(b) \leq a$ . Moreover  $\theta_a(\mu(b)) = \theta_a(b) - 1 \geq 2$ , so  $(a, \mu(b)) \notin \mathcal{B}$ , thus  $\mu(b) \neq X_1$ ,  $\lambda(\mu(b))$  makes sense and  $a < \lambda(\mu(b))$ . Then  $\lambda(\mu(b)) \neq X_0$  so  $\lambda^2(\mu(b))$  makes sense. Since  $b \in \mathcal{B}$ ,  $\lambda(\mu(b)) \leq \lambda(b)$  thus, using  $\lambda(b) < (a, \lambda(b))$  and Definition 8.2 and  $\lambda^2(\mu(b)) \leq \lambda^2(b) \leq a$ . We deduce that  $(a, \lambda(\mu(b))) \in \mathcal{B}$ .

Since  $\lambda(b) < (a, \lambda(b))$ , by Lemma 8.5,  $\lambda(b) = \text{ad}_a^m(v)$  for some  $m \in \mathbb{N}$  with  $v \neq X_1$  and  $\lambda(v) < a$ . By contradiction, if  $(a, \lambda(\mu(b))) < \lambda(\mu(b))$ , then, by Lemma 8.5,  $\lambda(\mu(b)) = \text{ad}_a^n(X_1)$  with  $n \in \mathbb{N}$ . Then, by Lemma 8.5,  $\lambda(\mu(b)) > \lambda(b)$ , which contradicts the assumption  $b \in \mathcal{B}$ .  $\square$

We now improve Proposition 8.14 by weakening the assumption and adding the slight precision reminiscent from the initialization of Lemma 8.25.

**Proposition 8.27.** *Let  $a < b \in \mathcal{B}$  such that  $(a, b) \notin \mathcal{B}$ ,  $(a, \lambda(b)) \in \mathcal{B}$  and  $\lambda(b) < (a, \lambda(b))$ . Then the estimates (8.11) still hold and, moreover, there exists  $c \in \text{supp}[a, b]$  such that  $\theta_{\lambda(b)}(c) = 2$  and  $\lambda(c) \leq (a, \lambda(b))$  (except when  $\theta_a(b) = 2$  and  $N_2(a, b) = 0$ ).*

*Proof.* The proof is the same as the one of Proposition 8.14 (by induction on  $\theta_a(b) \geq 2$ ). The initialization for  $\theta_a(b) = 2$  is performed in Lemma 8.25. The key point is that Lemma 8.26 allows to apply the induction hypothesis to the bracket  $[a, \mu(b)]$ . When  $\theta_a(b) > 2$ , the additional assertion on  $\text{supp}[a, b]$  comes from the fact that the first term of (8.12), named  $c_1$  in the proof of Proposition 8.14, satisfies  $\lambda(c_1) = (a, \lambda(b))$ .  $\square$

When  $(a, \lambda(b)) < \lambda(b)$ , we give induction relations for  $\|[a, b]\|_{\mathcal{B}}$  and  $N_2(a, b)$  based on the quantities  $\|[a, \mu(b)]\|_{\mathcal{B}}$  and  $N_2(a, \mu(b))$ .

**Proposition 8.28.** *Let  $a < b \in \mathcal{B}$  with  $\theta_a(b) \geq 3$ ,  $\lambda^2(b) \leq a$  and  $(a, \lambda(b)) < \lambda(b)$ .*

- If  $\lambda(\mu(b)) \leq (a, \lambda(b))$ , then  $N_2(a, b) \leq 2(N_2(a, \mu(b)) - P)$  and  $\|[a, b]\|_{\mathcal{B}} \leq 1 + \|[a, \mu(b)]\|_{\mathcal{B}} + N_2(a, \mu(b)) - P$
- Otherwise  $N_2(a, b) \leq 2^{\theta_a(\mu(b))-1} - 1 + 2(N_2(a, \mu(b)) - P)$  and  $\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta_a(\mu(b))-1} + \|[a, \mu(b)]\|_{\mathcal{B}} + N_2(a, \mu(b)) - P$ ,

where  $P$  is the fraction of  $\|[a, \mu(b)]\|_{\mathcal{B}}$  indexed by  $d \in \text{supp}[a, \mu(b)]$  such that  $\lambda(\mu(b)) < \lambda(d) \leq \lambda(b)$ .

*Proof.* The assumptions imply  $a < \lambda(b)$ ,  $a < \lambda(\mu(b))$ ,  $(a, \lambda(b)) \in \mathcal{B}$  and  $\theta_a(\mu(b)) = \theta_a(b) - 1 \geq 2$ . We study both terms of Jacobi's identity (8.12).

*First term*  $[(a, \lambda(b)), \mu(b)]$ . We have  $(a, \lambda(b)) < \mu(b)$  by Definition 8.2 because  $a < \lambda(\mu(b))$ .

- If  $\lambda(\mu(b)) \leq (a, \lambda(b))$  then  $c := ((a, \lambda(b)), \mu(b)) \in \mathcal{B}$  and  $\theta_{\lambda(b)}(c) = 1$  because  $\lambda(c) < \lambda(b)$ .
- If  $(a, \lambda(b)) < \lambda(\mu(b))$ , we apply Proposition 8.27 to the bracket  $[(a, \lambda(b)), \mu(b)]$ . Indeed  $\lambda^2(\mu(b)) \leq \lambda^2(b) \leq a < (a, \lambda(b))$ , thus  $\lambda(\mu(b)) < ((a, \lambda(b)), \lambda(\mu(b)))$ . By Lemma 4.8,  $\theta_{(a, \lambda(b))}(\mu(b)) \leq \theta_a(\mu(b))$ . Therefore Proposition 8.27 gives  $\|[a, \lambda(b)), \mu(b)]\|_{\mathcal{B}} \leq 2^{\theta_a(\mu(b))-1}$ ,  $N_2((a, \lambda(b)), \mu(b)) \leq 2^{\theta_a(\mu(b))-1} - 1$ . Each  $c \in \text{supp}[(a, \lambda(b)), \mu(b)]$  satisfies  $\theta_{\lambda(b)}(c) \leq \theta_{\lambda(\mu(b))}(c)$  by Lemma 4.8.

Hence, if  $\lambda(\mu(b)) \leq (a, \lambda(b))$ , the first term does not contribute to  $N_2(a, b)$  and contributes at most 1 to  $\|[a, b]\|_{\mathcal{B}}$ . Otherwise, the first term contributes at most  $2^{\theta_a(\mu(b))-1} - 2$  to  $N_2(a, b)$  and at most  $2^{\theta_a(\mu(b))-1}$  to  $\|[a, b]\|_{\mathcal{B}}$ .

*Second term*  $[\lambda(b), [a, \mu(b)]]$ . By Theorem 8.9,  $[a, \mu(b)] = \sum \alpha_d E(d)$  where the sum ranges over  $d \in \mathcal{B}$  such that  $a < \lambda(d) < \mu(b)$ ,  $\lambda(\mu(b)) < d$  and  $\theta_{\lambda(\mu(b))}(d) \leq 2$  using Lemma 8.12. Let  $d \in \text{supp}[a, \mu(b)]$ . Then  $\lambda(b) < d$ , by Definition 8.2, because  $\lambda^2(b) \leq a < \lambda(d)$ .

- If  $\theta_{\lambda(\mu(b))}(d) = 1$  then  $\theta_{\lambda(b)}(d) = 1$  by Lemma 4.8, thus  $c := (\lambda(b), d) \in \mathcal{B}$  and  $\theta_{\lambda(b)}(c) = 1$ .
- If  $\theta_{\lambda(\mu(b))}(d) = 2$  then  $\theta_{\lambda(b)}(d) \in \{1, 2\}$  by Lemma 4.8,
  - if  $\theta_{\lambda(b)}(d) = 1$  i.e.  $d$  is an index of the sum defining  $P$ , then  $c := (\lambda(b), d) \in \mathcal{B}$  and  $\theta_{\lambda(b)}(c) = 1$ ,
  - if  $\theta_{\lambda(b)}(d) = 2$  then  $\|[\lambda(b), d]\|_{\mathcal{B}} \leq 2$ .

Hence, the second term contributes at most  $2(N_2(a, \mu(b)) - P)$  to  $N_2(a, b)$  and at most  $(N_1(a, \mu(b)) + P) + 2(N_2(a, \mu(b)) - P) = \|[a, \mu(b)]\|_{\mathcal{B}} + N_2(a, \mu(b)) - P$  (thanks to Lemma 8.12) to  $\|[a, b]\|_{\mathcal{B}}$ , which concludes the proof.  $\square$

**Lemma 8.29.** *Let  $a < b \in \mathcal{B}$  with  $\theta_a(b) \geq 3$ ,  $\lambda^2(b) \leq a$ ,  $(a, \lambda(b)) < \lambda(b)$  and  $\lambda(\mu(b)) < (a, \lambda(\mu(b)))$ . Then  $\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta-1} - 1$  and  $N_2(a, b) \leq 2^{\theta-1} - 4$  where  $\theta := \theta_a(b)$ .*

*Proof.* Since  $\lambda(\mu(b)) < (a, \lambda(\mu(b)))$ , the bracket  $[a, \mu(b)]$  satisfies the hypotheses of Proposition 8.27 and  $\theta_a(\mu(b)) = \theta - 1$ . Thus,  $\|[a, \mu(b)]\|_{\mathcal{B}} \leq 2^{\theta-2}$  and  $N_2(a, \mu(b)) \leq 2^{\theta-2} - 1$ . Moreover, there exists  $c \in \text{supp}[a, \mu(b)]$  such that  $\lambda(\mu(b)) < \lambda(c) \leq (a, \lambda(\mu(b))) \leq (a, \lambda(b)) < \lambda(b)$ , except when  $\theta = 3$  and  $N_2(a, \mu(b)) = 0$ . The hypothesis  $\lambda(\mu(b)) < (a, \lambda(\mu(b)))$  implies that  $\lambda(\mu(b)) < (a, \lambda(b))$ . Hence, we can apply the first item of Proposition 8.28.

*Case  $\theta > 3$  or  $N_2(a, \mu(b)) > 0$ .* Then  $P \geq 1$ . Thus  $N_2(a, b) \leq 2(2^{\theta-2} - 1 - 1) = 2^{\theta-1} - 4$  and  $\|[a, b]\|_{\mathcal{B}} \leq 1 - 1 + 2^{\theta-2} + 2^{\theta-2} - 1 = 2^{\theta-1} - 1$ .

*Case  $\theta = 3$  and  $N_2(a, \mu(b)) = 0$ .* Then  $N_2(a, b) \leq 2(0 - 0) = 0 = 2^{\theta-1} - 4$  and  $\|[a, b]\|_{\mathcal{B}} \leq 1 - 0 + 2^{\theta-2} + 0 = 3 = 2^{\theta-1} - 1$ .  $\square$

**Lemma 8.30.** *Let  $a < b \in \mathcal{B}$  with  $\theta_a(b) \geq 2$ ,  $\lambda^2(b) \leq a$  and  $(a, \lambda(b)) < \lambda(b)$ . Then*

$$\|[a, b]\|_{\mathcal{B}} \leq A(\theta) \quad \text{and} \quad N_2(a, b) \leq A(\theta) - \theta, \quad \text{where } \theta := \theta_a(b). \quad (8.34)$$

*Proof.* We proceed by induction on  $\theta := \theta_a(b) \geq 2$ .

*Initialization for  $\theta = 2$ .* According to Lemma 8.25,  $\|[a, b]\|_{\mathcal{B}} \leq 2 = A(2)$ ,  $N_2(a, b) \leq 1$ , and moreover, if  $N_2(a, b) = 1$ , the unique  $c \in \text{supp}[a, b]$  such that  $\theta_{\lambda(b)}(c) = 2$  satisfies  $\lambda(c) \leq (a, \lambda(b))$ . Thus, by assumption,  $\lambda(c) < \lambda(b)$  so  $\theta_{\lambda(b)}(c) = 1$ . Hence  $N_2(a, b) = 0 = A(2) - 2$ . So (8.34) holds.

*Induction for  $\theta \geq 3$ .* If  $\lambda(\mu(b)) < (a, \lambda(\mu(b)))$ , estimate (8.34) follows from Lemma 8.29 since  $2^{\theta-1} - 1 \leq A(\theta) - \theta$  and  $2^{\theta-1} - 1 \leq A(\theta)$  for  $\theta \geq 3$ . Otherwise, when  $(a, \lambda(\mu(b))) < \lambda(\mu(b))$ , the bracket  $[a, \mu(b)]$  satisfies the hypotheses of Lemma 8.30. Thus  $\|[a, \mu(b)]\|_{\mathcal{B}} \leq A(\theta - 1)$  and  $N_2(a, \mu(b)) \leq A(\theta - 1) - (\theta - 1)$ . Moreover, by Proposition 8.28, even if  $P = 0$ ,

$$\begin{aligned} N_2(a, b) &\leq 2^{\theta-2} - 1 + 2N_2(a, \mu(b)) \\ &\leq 2^{\theta-2} - 1 + 2(\theta - 4)2^{\theta-3} + 2 = A(\theta) - \theta \end{aligned} \quad (8.35)$$

and

$$\begin{aligned} \|[a, b]\|_{\mathcal{B}} &\leq 2^{\theta-2} + \|[a, \mu(b)]\|_{\mathcal{B}} + N_2(a, \mu(b)) \\ &\leq 2^{\theta-2} + A(\theta - 1) + A(\theta - 1) - (\theta - 1) = A(\theta), \end{aligned} \quad (8.36)$$

so (8.34) is proved.  $\square$

**Theorem 8.31.** *Let  $a < b \in \mathcal{B}$  with  $(a, b) \notin \mathcal{B}$  such that  $\lambda^2(b) \leq a$ . Then  $\|[a, b]\|_{\mathcal{B}} \leq A(\theta_a(b))$ .*

*Proof.* This follows from Proposition 8.28 when  $\lambda(b) < (a, \lambda(b))$  and from Lemma 8.30 when  $(a, \lambda(b)) < \lambda(b)$  (the situation  $\lambda(b) = (a, \lambda(b))$  being impossible, since  $|(a, \lambda(b))| > |\lambda(b)|$ ).  $\square$

We eventually study the following particular subcase, which provides a theoretical proof to a numerical claim made in the proof of Lemma 8.22.

**Corollary 8.32.** *Let  $\theta \geq 2$ . Then  $\|[X_0, \text{ad}_{(X_0, X_1)}^{\theta-1}(X_1)]\|_{\mathcal{B}} \leq A(\theta) - 1$ .*

*Proof.* Let  $a := X_0$  and  $b_{\theta} := \text{ad}_{(X_0, X_1)}^{\theta-1}(X_1)$ . When  $\theta = 2$ , by Jacobi's identity,

$$[a, b_2] = [[X_0, [X_0, X_1]], X_1] + [[X_0, X_1], [X_0, X_1]] = [[X_0, [X_0, X_1]], X_1], \quad (8.37)$$

where  $((X_0, (X_0, X_1)), X_1) \in \mathcal{B}$ . Hence  $\|[a, b_2]\|_{\mathcal{B}} = 1 = A(2) - 1$  and  $N_2(a, b_2) = 0$ , which, thanks to the cancellation in (8.37), are strictly better estimates than in Lemma 8.25.

Moreover, for every  $\theta \geq 2$ ,  $\lambda^2(b_{\theta}) = X_0 = a$  and  $(a, \lambda(b)) = (X_0, (X_0, X_1)) < (X_0, X_1) = \lambda(b)$ . Since  $\mu(b_{\theta}) = b_{\theta-1}$ , the estimate  $\|[a, b_{\theta}]\|_{\mathcal{B}} \leq A(\theta) - 1$  is propagated by induction by (8.36).  $\square$

#### 8.8.4 Sharp bound in the general case

We prove Theorem 8.37 which extends Theorem 8.31 to all brackets. We start with a definition and preliminary results concerning situations where the structure constants are much smaller than expected (see Lemma 8.36) and is related to Lemma 8.17.

**Definition 8.33.** *We define a relation  $\ll$  on  $\mathcal{B}$  by the following conditions: for  $a_0, a \in \mathcal{B}$ ,  $a_0 \ll a$  when  $a_0 \neq X_1$ ,  $a \neq X_0$ , and either  $a = X_1$ , or  $a = (a_0, X_1)$ , or  $a_0 < \lambda(a)$ .*

**Lemma 8.34.** *Let  $a_0 \ll a \in \mathcal{B}$ . Then  $a_0 < a$ . If  $b \in \mathcal{B} \setminus X$  is such that  $a < b$ , then  $a_0 < \lambda(b)$ .*

*Proof.* The first item follows from the relations  $a_0 < (a_0, X_1)$  and  $\lambda(a) < a$ . For the second item, let  $b \in \mathcal{B} \setminus X$  such that  $a < b$ . Then  $a \neq X_1$ . If  $a = (a_0, X_1)$  then the relation  $a < b$  implies  $a_0 < \lambda(b)$  because  $\mu(b) < X_1$ . If  $a_0 < \lambda(a)$  then  $a_0 < \lambda(b)$  because  $\lambda(a) \leq \lambda(b)$ .  $\square$

**Remark 8.35.** *Lemma 8.34 implies that the relation  $\ll$  is transitive. In fact, it is an order on  $\mathcal{B}$ . Indeed, the conditions  $a \ll b$  and  $b \ll a$  are incompatible since each of them implies the same comparison for the relation  $<$ . However, since  $(X_0, X_1)$  and  $(X_0, (X_0, X_1))$  cannot be compared by  $\ll$ , this order is not total.*

**Lemma 8.36.** *Let  $a_0 \ll a < b < X_1 \in \mathcal{B}$ . Then  $\theta_{a_0}(b) \geq 2$  and  $\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta_{a_0}(b)-2}$ .*

*Proof.* Since  $a_0 \ll a < b$ , Lemma 8.34 shows that  $a_0 < \lambda(b)$ , so  $\theta_{a_0}(b) \geq 2$ . We prove the estimate on  $\|[a, b]\|_{\mathcal{B}}$  by induction on  $\theta_{a_0}(b) \geq 2$ .

*Initialization for  $\theta_{a_0}(b) = 2$ .* Then  $\lambda^2(b) \leq a_0$  so  $\lambda(b) \leq a$  (indeed, either  $a = (a_0, X_1)$  and then  $\lambda(b) = (\lambda^2(b), \mu\lambda(b)) \leq (a_0, X_1) = a$  or  $a_0 < \lambda(a)$  and then  $\lambda^2(b) < \lambda(a)$  thus  $\lambda(b) < a$ ). Therefore,  $(a, b) \in \mathcal{B}$  and  $\|[a, b]\|_{\mathcal{B}} = 1 = 2^{\theta_{a_0}(b)-2}$ .

*Inductive step for  $\theta_{a_0}(b) \geq 3$ .* Since  $a < \lambda(b)$ ,  $a_0 < \lambda^2(b)$  by Lemma 8.34 and  $\theta_{a_0}(\lambda(b)) \geq 2$ . Moreover, either  $\mu(b) = X_1$ , or the same argument implies that  $\theta_{a_0}(\mu(b)) \geq 2$ .

- *Case  $\mu(b) = X_1$ .* In this case,  $[a, b] = [[a, \lambda(b)], X_1] + [\lambda(b), (a, X_1)]$ . For the first term, by the induction hypothesis,  $\|[a, \lambda(b)], X_1]\|_{\mathcal{B}} \leq 2^{\theta_{a_0}(\lambda(b))-2} = 2^{\theta_{a_0}(b)-3}$ . For the second term, if  $\lambda(b) < (a, X_1)$ , then  $(\lambda(b), (a, X_1)) \in \mathcal{B}$  so  $\|[\lambda(b), (a, X_1)]\|_{\mathcal{B}} = 1 \leq 2^{\theta_{a_0}(b)-3}$ . If  $(a, X_1) < \lambda(b)$ , then  $a_0 \ll a \ll (a, X_1)$ , so by the induction hypothesis,  $\|[(a, X_1), \lambda(b)]\|_{\mathcal{B}} \leq 2^{\theta_{a_0}(\lambda(b))-2} = 2^{\theta_{a_0}(b)-3}$ . Therefore, summing the estimates proves that  $\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta_{a_0}(b)-2}$ .

- *Case  $\mu(b) < X_1$ .* We decompose  $[a, b]$  using Jacobi's identity.

By the induction hypothesis,  $[a, \lambda(b)] = \sum_d \alpha_d E(d)$  with  $\sum_d |\alpha_d| \leq 2^{\theta_{a_0}(\lambda(b)) - 2}$ . Let  $d \in \text{supp}[a, \lambda(b)]$ . If  $\mu(b) < d$ , the inequality  $\lambda(d) < \lambda(b) < \mu(b)$  from Theorem 8.9 shows that  $(\mu(b), d) \in \mathcal{B}$ . If  $d < \mu(b)$ , the inequality  $\lambda(d) \geq a > a_0$  shows that, by the induction hypothesis,  $\|[d, \mu(b)]\|_{\mathcal{B}} \leq 2^{\theta_{a_0}(\mu(b)) - 2}$ . Therefore  $\|[a, \lambda(b)], \mu(b)\|_{\mathcal{B}} \leq 2^{\theta_{a_0}(b) - 4}$ .

Since  $\mu(b) \neq X_1$ ,  $\lambda(\mu(b)) \geq \lambda(\lambda(b)) > a_0$ . By the induction hypothesis,  $[a, \mu(b)] = \sum_d \alpha_d E(d)$  with  $\sum_d |\alpha_d| \leq 2^{\theta_{a_0}(\mu(b)) - 2}$ . Let  $d \in \text{supp}[a, \mu(b)]$ . If  $d < \lambda(b)$ , then using the induction hypothesis again shows that  $\|[d, \lambda(b)]\|_{\mathcal{B}} \leq 2^{\theta_{a_0}(\lambda(b)) - 2}$ . If  $\lambda(b) < d$ , then  $\lambda(\mu(b)) \leq \lambda(b) < d$  thus, by Theorem 8.9 and Lemma 4.8,  $\theta_{\lambda(b)}(d) \leq \theta_{\lambda(\mu(b))}(d) \leq 2$  thus  $\|[\lambda(b), d]\|_{\mathcal{B}} \leq 2$ . Taking into account that  $\theta_{a_0}(\lambda(b)) \geq 2$  we conclude that  $\|[\lambda(b), [a, \mu(b)]]\|_{\mathcal{B}} \leq 2^{\theta_{a_0}(b) - 3}$ .

Combining both parts and using Jacobi's identity, one gets  $\|[a, b]\|_{\mathcal{B}} \leq 2^{\theta_{a_0}(b) - 2}$ .  $\square$

**Theorem 8.37.** *Let  $a < b \in \mathcal{B}$ . Then  $\|[a, b]\|_{\mathcal{B}} \leq A(\theta_a(b))$ . Moreover, if  $b \neq X_1$  and  $(a, \lambda(b)) \notin \mathcal{B}$ , then the inequality is strict.*

*Proof.* We proceed by induction on  $\theta_a(b)$ . We may assume that  $(a, b) \notin \mathcal{B}$ . Thanks to Theorem 8.31, which corresponds to the case  $\theta_a(\lambda(b)) = 1$ , we may further assume that  $\theta_a(\lambda(b)) \geq 2$ , i.e. that  $\lambda(b) \neq X_0$  and  $a < \lambda^2(b)$ .

*Case  $\mu(b) = X_1$ .* In this case,  $[a, b] = [[a, \lambda(b)], X_1] - [(a, X_1), \lambda(b)]$ . By the induction hypothesis, we have  $\|[a, \lambda(b)]\|_{\mathcal{B}} \leq A(\theta_a(\lambda(b)))$ , and by Lemma 8.36 (since  $(a, X_1) < \lambda(b) < X_1$ ), we have  $\|[(a, X_1), \lambda(b)]\|_{\mathcal{B}} \leq 2^{\theta_a(\lambda(b)) - 2}$ . Combining both estimates gives:

$$\|[a, b]\|_{\mathcal{B}} \leq A(\theta_a(b) - 1) + 2^{\theta_a(b) - 3} = A(\theta_a(b)) - ((\theta - 2)2^{\theta - 2} + 1) < A(\theta_a(b)). \quad (8.38)$$

*Case  $\mu(b) \neq X_1$ .* Using Jacobi's identity, we write  $[a, b] = [[a, \lambda(b)], \mu(b)] + [\lambda(b), [a, \mu(b)]]$ . We may assume that  $\theta_a(\mu(b)) \geq 2$  (otherwise,  $\lambda(\mu(b)) \leq a$  implies that  $\lambda^2(b) \leq a$  so  $\theta_a(\lambda(b)) = 1$ ).

- For the first term, by the induction hypothesis and Theorem 8.9,  $[a, \lambda(b)] = \sum_d \alpha_d E(d)$ , with  $a < \lambda(d) < \lambda(b)$ , and  $\sum_d |\alpha_d| \leq A(\theta_a(\lambda(b)))$ . Let  $d \in \text{supp}[a, \lambda(b)]$ . If  $\mu(b) < d$ , then  $(\mu(b), d) \in \mathcal{B}$  and  $\|[\mu(b), d]\|_{\mathcal{B}} = 1$ . If  $d < \mu(b)$ , then by Lemma 8.36 (since  $\mu(b) \neq X_1$  and  $a \ll d$ ), we have  $\|[d, \mu(b)]\|_{\mathcal{B}} \leq 2^{\theta_a(\mu(b)) - 2}$ . Therefore,  $\|[a, \lambda(b)], \mu(b)\|_{\mathcal{B}} \leq A(\theta_a(\lambda(b)))2^{\theta_a(\mu(b)) - 2}$ .
- For the second term, by the induction hypothesis and Theorem 8.9,  $[a, \mu(b)] = \sum_d \alpha_d E(d)$ , with  $a < \lambda(d) < \mu(b)$ , and  $\sum_d |\alpha_d| \leq A(\theta_a(\mu(b)))$ . Moreover, we know that either  $d \leq \lambda(\mu(b)) \leq \lambda(b)$ , or  $\theta_{\lambda(b)}(d) \leq \theta_{\lambda(\mu(b))}(d) \leq 2$ . If  $\lambda(b) < d$ , then  $\|[\lambda(b), d]\|_{\mathcal{B}} \leq 2$ . If  $d < \lambda(b)$ , by Lemma 8.36,  $\|[d, \lambda(b)]\|_{\mathcal{B}} \leq 2^{\theta_a(\lambda(b)) - 2}$ . In both cases  $\|[d, \lambda(b)]\|_{\mathcal{B}} \leq 2^{\theta_a(\lambda(b)) - 1}$ , and thus  $\|[\lambda(b), [a, \mu(b)]]\|_{\mathcal{B}} \leq A(\theta_a(\mu(b)))2^{\theta_a(\lambda(b)) - 1}$ .

Finally, we get the estimate:

$$\|[a, b]\|_{\mathcal{B}} \leq A(\theta_a(\lambda(b)))2^{\theta_a(\mu(b)) - 2} + A(\theta_a(\mu(b)))2^{\theta_a(\lambda(b)) - 1}. \quad (8.39)$$

Since  $\theta_a(\lambda(b)) \geq 1$  and  $\theta_a(\mu(b)) \geq 1$ , the result then comes from following claim:

$$\forall p, q \geq 1, \quad A(p)2^{q-2} + A(q)2^{p-1} < A(p)2^{q-1} + A(q)2^{p-1} \leq A(p+q). \quad (8.40)$$

The first inequality is clear since  $A(p) \neq 0$ . For the second inequality, dividing by  $2^{p+q-3}$ , and using formula that defines  $A$  shows that (8.40) is equivalent to:

$$p+q-6 + \frac{p+1}{2^{p-2}} + \frac{q+1}{2^{q-2}} \leq 2(p+q-3) + \frac{p+q+1}{2^{p+q-3}}. \quad (8.41)$$

This inequality can be checked directly when  $p \in \{1, 2\}$  (and thus holds when  $q \in \{1, 2\}$  by symmetry), and follows from  $\frac{p+1}{2^{p-2}} + \frac{q+1}{2^{q-2}} \leq \frac{p+q+2}{2} \leq p+q$  when  $p, q \geq 3$ . Therefore (8.40) holds, which concludes the proof.  $\square$

## 9 A super-geometric Hall set

When  $X$  is infinite, we know from Corollary 4.15 that the structure constants can grow super-geometrically with respect to the length of the involved brackets and to  $\theta_a(b)$ . The main goal of this section is to prove Theorem 1.8, which illustrates that, even in the most constrained case when  $|X| = 2$ , there exists a Hall set on  $X$  whose structure constants grow super-geometrically. We conclude that the structure constants can grow super-geometrically both with respect to the length of the brackets and to  $\theta_a(b)$ . Hence, the results obtained in Section 6 (for the length-compatible Hall sets) and Section 7 (for the Lyndon sets), where the structure constants grow geometrically with the length of the involved brackets, cannot be extended to all Hall sets, even when only two indeterminates are considered.

We start by describing a natural strategy in Section 9.1 which however fails to prove the desired conclusion. We then construct a specific order and the associated Hall set in Section 9.2. We compute the exact decomposition in the basis of a nasty Lie bracket in Section 9.3. Eventually, we prove a refined version of Theorem 1.8 in Section 9.4 by optimizing the choice of the nasty bracket.

### 9.1 A natural but unfitted strategy

Assume that  $X$  is finite. A possible strategy to tackle Theorem 1.8 would be to attempt to realize in  $\text{Br}(X)$  the critical construction of Corollary 4.15.

One could attempt to construct a Hall set  $\mathcal{B} \subset \text{Br}(X)$ , and, for every  $p$  large enough, a free alphabetic subset  $X_p := \{x_0, x_1, \dots, x_p\} \subset \mathcal{B}$ . Then, using the isometry described in Proposition 3.5, one could rely on the construction of Proposition 4.14 to obtain  $\|[x_0, b]\|_{\mathcal{B}} = \lfloor e(p-1)! \rfloor$ , where  $b := (\dots((x_1, x_p), x_{p-1}), \dots, x_2)$  (proving that  $b \in \mathcal{B}$  and achieving the equality would require that the order on  $\mathcal{B} \cap \text{Br}_{X_p}$  has been chosen as described in Lemma 4.13).

Let us explain why this strategy fails to obtain a super-geometric lower growth with respect to the length of the considered brackets.

Let  $\mathcal{B} \subset \text{Br}(X)$  be a Hall set. The construction would require in particular that  $x_1, \dots, x_p$  are  $p \geq 1$  distinct elements of  $\mathcal{B}$ . Let us derive a lower bound for  $|b| = |x_1| + \dots + |x_p|$  (not even considering the fact that executing the strategy would also require additional constraints, such as  $X_p$  being a free alphabetic subset of  $\mathcal{B}$ ). In particular  $|b| \geq |y_1| + \dots + |y_p| =: C_p$  where  $y_1, \dots, y_p$  are the first  $p$  elements of  $\mathcal{B}$ , sorted by non-decreasing length (breaking ties using for example the order on  $\mathcal{B}$ ).

For any  $\ell \in \mathbb{N}^*$ ,

$$|\{t \in \mathcal{B}; |t| = \ell\}| \leq |X|^\ell. \quad (9.1)$$

From Witt's formula [50], the inequality is strict for every  $\ell > 1$ . However, this increases the lower-bound  $C_p$  since there are fewer brackets of small length. Hence, we continue the discussion as if (9.1) was an equality for all  $\ell \in \mathbb{N}^*$ .

Let  $L := \max |y_i| = |y_p|$ . Then  $|X|^{L+1} \geq p+1$ . And, since the  $y_i$  were chosen by non-decreasing length,

$$C_p = \sum_{i=1}^p |y_i| \geq \sum_{\ell=1}^{L-1} \ell |X|^\ell \geq (L-1) |X|^{L-1} \geq \frac{1}{|X|^2} (p+1) \left( \frac{\ln(p+1)}{\ln |X|} - 2 \right). \quad (9.2)$$

Thus, for  $p$  large enough, one has  $p \ln p \leq (2|X|^2 \ln |X|) |b|$ .

Even if the construction is performed such that  $\|[x_0, b]\|_{\mathcal{B}} = \lfloor e(p-1)! \rfloor$ , the bounds associated with Stirling's approximation (see e.g. [34]) yield, for  $p \geq 3$ ,

$$\|[x_0, b]\|_{\mathcal{B}} \leq p^p = e^{p \ln p} \leq |X|^{2|X|^2 |b|}, \quad (9.3)$$



which is a geometric bound with respect to  $|b|$ .

Hence, we develop in the sequel a method which avoids this path. In particular, the brackets and the order we consider are not the same as in Corollary 4.15.

## 9.2 Construction of an appropriate Hall set

Let  $X_0, X_1 \in X$ . We start by introducing the elementary building blocks of our construction. For  $i \in \mathbb{N}$ , define

$$A_i := \text{ad}_{X_0}^i(X_1). \quad (9.4)$$

**Lemma 9.1.** *The following subset of  $\text{Br}(X)$  is free in the sense of Lemma 1.19:*

$$A := \{A_i; i \in \mathbb{N}\}. \quad (9.5)$$

*Proof.* All elements of  $A$  contain  $X_1$  exactly once, while elements of  $(\text{Br}_A, \text{Br}_A)$  contain  $X_1$  at least twice, so  $A \cap (\text{Br}_A, \text{Br}_A) = \emptyset$ .  $\square$

We now introduce a subset of  $\text{Br}(X)$  within which we will be working. Let

$$G := \{X_0, X_1\} \cup G^* \quad \text{where} \quad G^* := \{b \in \text{Br}_A; X_1 \notin \Lambda(b)\}, \quad (9.6)$$

where  $\Lambda(b)$  is the set of iterated left factors of  $b$  (see (1.10)).

**Lemma 9.2.**  *$G$  is  $\lambda$ -stable and  $X_1 \notin G^*$ .*

*Proof.*  $X_1 \notin G^*$  because  $X_1 \in \Lambda(X_1) = \{\lambda^0(X_1)\} = \{X_1\}$ . Moreover, if  $b \in G^*$ , either  $b = A_i$  for some  $i \geq 1$  so that  $\lambda(b) = X_0 \in G$ , or  $b = (b_1, b_2)$  with  $b_1, b_2 \in \text{Br}_A$  and  $X_1 \notin \Lambda(b) = \Lambda(b_1) \cup \{b\}$ . Hence  $\lambda(b) = b_1 \in G^*$ . Hence  $\lambda(G \setminus X) = \lambda(G^*) \subset G^* \cup \{X_0\} \subset G$ .  $\square$

**Definition 9.3** (Score on  $\text{Br}_A$ ). *We define a score map  $s : \text{Br}(A) \rightarrow \mathbb{N}$ . For  $i \in \mathbb{N}$ , set*

$$s(A_i) := \begin{cases} i & \text{if } i \in \{0, 1, 2\}, \\ 3 \cdot 2^{i-3} & \text{if } i \geq 3. \end{cases} \quad (9.7)$$

*By induction, if  $b \in \text{Br}(A)$  is of the form  $\langle b_1, b_2 \rangle$  with  $b_1, b_2 \in \text{Br}(A)$ , we set*

$$s(\langle b_1, b_2 \rangle) := s(b_1) + s(b_2). \quad (9.8)$$

*In particular, for all  $b \in \text{Br}_A \setminus \{X_1\}$ ,  $s(b) > 0$ . Then, since  $A$  is free, by Lemma 1.19,  $\text{Br}_A$  is isomorphic to  $\text{Br}(A)$ . Thus, setting  $s(\mathbf{1}(b)) := s(b)$  for all  $b \in \text{Br}(A)$  extends  $s$  to  $\text{Br}_A$ .*

**Definition 9.4** (Iteration by  $X_1$  and germs). *For  $b \in \text{Br}(X)$  and  $\nu \in \mathbb{N}$ , we will use in the sequel of this section the notation  $b1^\nu$  instead of  $\text{ad}_{X_1}^\nu(b)$  for concision. For every  $b \in G^*$ , there exists a unique couple  $(b^*, \nu(b)) \in G^* \times \mathbb{N}$  such that  $b = b^*1^{\nu(b)}$  and  $\mu(b^*) \neq X_1$ . One checks that  $b^* \in G^*$  because  $X_1 \notin \Lambda(b) \supset \Lambda(b^*)$ .*

**Definition 9.5** (Order on  $G$ ). *We define an order on  $G$  by setting  $X_0 < G^* < X_1$  and, inside  $G^*$ , the lexicographic order on the quadruple  $s(b), \lambda(b^*), \mu(b^*), \nu(b)$ .*

**Proposition 9.6.** *There exists a Hall order on  $\text{Br}(X)$  extending the above order on  $G$ .*

*Proof.* Since  $G$  is  $\lambda$ -stable, by Proposition 1.33, it is sufficient to check that the order defined in Definition 9.5 is a Hall order on  $G$ . First, this order is total on  $G$ . Indeed, the triple  $\lambda(b^*), \mu(b^*), \nu(b)$  defines  $b$  uniquely. Second, this order is compatible with  $\lambda$ . Indeed, let  $b \in G^*$ . If  $b = A_i$  for some  $i \geq 1$ ,  $\lambda(b) = X_0 < b$ . Otherwise  $b = (b_1, b_2)$  with  $b_1 \in G^*$  and  $b_2 \in \text{Br}_A$ . If  $b_2 = X_1$ , then  $s(b) = s(b_1)$  by (9.8) and  $b^* = b_1^*$  so  $\lambda(b^*) = \lambda(b_1^*)$  and  $\mu(b^*) = \mu(b_1^*)$ , and  $\nu(b) = \nu(b_1) + 1$ . So  $b_1 < b$ . If  $b_2 \neq X_1$ ,  $s(b_2) > 0$  and, by (9.8),  $s(b_1) < s(b)$  which implies that  $b_1 < b$ .  $\square$

In the sequel, we consider  $\mathcal{B} \subset \text{Br}(X)$  the Hall set associated with any such order on  $\text{Br}(X)$ .

### 9.3 Decomposition of a nasty Lie bracket

Let  $\nu \geq 0$ . We then consider the sequence of brackets defined by induction as

$$B_2^\nu := A_2 1^\nu \quad \text{and} \quad B_{k+1}^\nu := (B_k^\nu, A_{k+1}) \quad \text{for } k \geq 2. \quad (9.9)$$

We prove the following equality.

**Proposition 9.7.** *Let  $p \geq 2$  and  $\nu \geq 0$ . Then  $A_1$  and  $B_p^\nu$  belong to  $\mathcal{B}$  and*

$$\|[A_1, B_p^\nu]\|_{\mathcal{B}} = p^\nu + p - 2. \quad (9.10)$$

*Proof. Step 1: We check that the considered brackets are indeed in  $\mathcal{B}$ .* By definition of the order,  $X_0 < X_1$ , so, for all  $i \in \mathbb{N}$ ,  $A_i \in \mathcal{B}$ . Since  $G^* < X_1$ , for each  $g \in G^*$  and  $\nu \in \mathbb{N}$ ,  $g 1^\nu \in \mathcal{B}$ . In particular,  $B_2^\nu = A_2 1^\nu \in \mathcal{B}$ . Moreover, for all  $k \geq 2$ ,  $B_k^\nu \in G^*$  and, by induction, one checks that  $s(B_k^\nu) = 2 + \sum_{i=3}^k 3 \cdot 2^{i-3} = 3 \cdot 2^{k-2} - 1$  and  $s(A_{k+1}) = 3 \cdot 2^{k-2}$ . Thus,  $B_k^\nu < A_{k+1}$  and  $\lambda(A_{k+1}) = X_0 < B_k^\nu$ . This guarantees that  $B_k^\nu \in \mathcal{B}$  for all  $\nu \geq 0$  and  $k \in \mathbb{N}$ .

*Step 2: We decompose the bracket on the basis.* Applying the iterated Jacobi identity to  $[A_1, B_p^\nu]$  yields

$$[A_1, B_p^\nu] = [\dots [[A_1, B_2^\nu], A_3], \dots], A_p] + \sum_{k=2}^{p-1} [[B_k^\nu, [A_1, A_{k+1}]], \dots], A_p]. \quad (9.11)$$

First, we remark that for each  $k \in \llbracket 2, p-1 \rrbracket$ , the element in the sum belongs to  $\pm \mathcal{E}(\mathcal{B})$ . Indeed,  $(A_1, A_{k+1}) \in \mathcal{B}$  because  $s(A_1) < s(A_{k+1})$  and  $\lambda(A_{k+1}) = X_0 < A_1$ . Then,  $(B_k^\nu, (A_1, A_{k+1})) \in \mathcal{B}$  because  $s(B_k^\nu) = 3 \cdot 2^{k-2} - 1 < 3 \cdot 2^{k-2} + 1 = s((A_1, A_{k+1}))$  and  $\lambda((A_1, A_{k+1})) = A_1 < B_k^\nu$ . Let  $C_{k+1} := (B_k^\nu, (A_1, A_{k+1}))$ . We prove by induction on  $\ell \in \llbracket k+1, p \rrbracket$  that  $C_\ell \in \mathcal{B}$  where  $C_{\ell+1} := (A_\ell, C_\ell)$ . Indeed, one checks that  $s(C_\ell) = 3 \cdot 2^{\ell-2} = s(A_\ell)$  so that  $A_\ell < C_\ell$  because  $\lambda(A_\ell^*) = \lambda(A_\ell) = X_0 < \lambda(C_\ell)$ . Moreover,  $\lambda(C_\ell) < A_\ell$  because  $\lambda(C_\ell) = A_{\ell-1}$  if  $\ell \geq k+2$  and  $\lambda(C_{k+1}) = B_k^\nu$ . Hence the sum of  $p-2$  terms is already expressed on  $\mathcal{B}$ .

Second, we take care of the first term of the right-hand side of (9.11). By Lemma B.3 (applied to the derivation  $\underline{\text{ad}}_{X_1}$  on  $\mathcal{L}(X)$ ,  $k \leftarrow p$ ,  $b_k \leftarrow A_2$ ,  $b_{k-1} \leftarrow A_1$  and  $b_i \leftarrow A_{p+1-i}$  for  $1 \leq i \leq k-2$ ),

$$\begin{aligned} & (-1)^{p-2} [\dots [[A_1, B_2^\nu], A_3], \dots], A_p] \\ &= \sum_{j_1 + \dots + j_p = \nu} (-1)^{\nu - j_p} \binom{\nu}{j_1, \dots, j_p} [A_p 1^{j_1}, [A_{p-1} 1^{j_2}, [\dots, [A_1 1^{j_{p-1}}, A_2] \dots]]] 1^{j_p}. \end{aligned} \quad (9.12)$$

One checks that all these brackets lie in  $\mathcal{B}$ . First,  $s(A_1) < s(A_2)$  so  $A_1 1^{j_{p-1}} < A_2$  and  $\lambda(A_2) = X_0 < A_1 1^{j_{p-1}}$ , hence  $(A_1 1^{j_{p-1}}, A_2) \in \mathcal{B}$ . Then, by construction, for  $k \geq 3$ ,  $s(A_k) = 3 \cdot 2^{k-3} = \sum_{i=1}^{k-1} s(A_i)$  but  $\lambda((A_k 1^{j_{1+p-k}})^*) = \lambda(A_k) = X_0 < \lambda((A_{k-1} 1^{j_{2+p-k}}, \dots)^*) = A_{k-1}$  so the left members are lower than the right members. Eventually since  $s(A_i)$  is increasing, these brackets also satisfy the Hall condition that the left part of the right member is lower than the left member.

Hence, the summation formula (B.2) yields

$$\|[\dots [A_1, B_2^\nu], A_3], \dots], A_p]\|_{\mathcal{B}} = p^\nu. \quad (9.13)$$

This concludes the proof, because all considered brackets of  $\mathcal{B}$  are distinct.  $\square$

### 9.4 Conclusion of the proof by optimization

We now prove the following result, which of course implies its unquantified counterpart Theorem 1.8 stated in the introduction. In Remark 9.9, we then discuss the consequences of our construction for  $\theta$ -based lower bounds.

**Theorem 9.8.** *For the Hall set  $\mathcal{B} \subset \text{Br}(X)$  constructed in Section 9.2, there exists  $c_0 > 0$  such that, for every  $n \in \mathbb{N}^*$  large enough, there exist  $a < b \in \mathcal{B}$  with  $|a| = 2$  and  $|b| = n$  such that*

$$\|[a, b]\|_{\mathcal{B}} \geq \left( \sqrt{\frac{n}{c_0 \ln n}} \right)^n. \quad (9.14)$$

*Proof.* Let  $n \geq 4$ . As in the previous paragraph, we consider  $a := A_1$  and  $b := B_p^\nu$ . We now optimize the choice of  $\nu \in \mathbb{N}$  and  $p \geq 2$  to obtain the claimed estimate. For  $i \in \mathbb{N}$ , we have  $|A_i| = i + 1$ . Moreover, for  $\nu \in \mathbb{N}$  and  $p \geq 2$ ,

$$|B_p^\nu| = |B_2^\nu| + \sum_{i=3}^p |A_i| = \nu + 3 + \sum_{i=3}^p (1 + i) = \nu + \frac{(p+1)(p+2)}{2} - 3. \quad (9.15)$$

For  $p \geq 2$  such that  $p^2 \leq n$ , we define  $\nu_p := n + 3 - \frac{(p+1)(p+2)}{2} \geq 0$ . From now on  $\nu := \nu_p$ , which ensures that  $|b| = n$ . Moreover,  $\nu_p \geq n - p^2$ . By Proposition 9.7,

$$\|[a, b]\|_{\mathcal{B}} \geq p^{\nu_p} \geq p^{n-p^2}. \quad (9.16)$$

We must now choose  $p \in \llbracket 2, \lfloor \sqrt{n} \rrbracket \rrbracket$  to take advantage of this lower bound. We set

$$p := \left\lfloor \sqrt{\frac{n}{\ln n}} \right\rfloor. \quad (9.17)$$

For  $n$  large enough,  $p \geq 2$  and the construction is valid. Moreover

$$n - p^2 \geq n - \frac{n}{\ln n} = n \left( 1 - \frac{1}{\ln n} \right). \quad (9.18)$$

Thus

$$\|[a, b]\|_{\mathcal{B}} \geq M(n)^n \quad (9.19)$$

where

$$M(n) := \left( \sqrt{\frac{n}{\ln n}} - 1 \right)^{1 - \frac{1}{\ln n}} \geq \sqrt{\frac{n}{c_0 \ln n}} \quad (9.20)$$

for some appropriate choice of  $c_0 > 0$  and  $n$  large enough using elementary asymptotic analysis, which concludes the proof.  $\square$

**Remark 9.9.** *Let  $a := A_1$  and  $b := B_p^\nu$  as above. Then  $\theta_a(b) = p - 1 + \nu$ . Indeed, for every  $i \geq 2$ ,  $a < A_i$  and  $(a, A_i) \in \mathcal{B}$ , and  $a < X_1$  and  $(a, X_1) \in \mathcal{B}$ . Hence, a similar procedure as the one above proves that, for every  $\theta \in \mathbb{N}$  large enough, there exists  $a, b \in \mathcal{B}$  with  $\theta_a(b) = \theta$  and*

$$\|[a, b]\|_{\mathcal{B}} \geq \frac{1}{\theta^2} \left( \frac{\theta}{e \ln \theta} \right)^\theta, \quad (9.21)$$

*which is therefore super-geometric (relative to  $\theta$ ) and qualitatively not very far from the general  $\theta$ -based upper bound (4.7). This lower bound can for example be obtained with the choice  $p := \lfloor \theta / \ln \theta \rfloor$  and  $\nu_p = \theta + 1 - p$  and then follows from elementary asymptotic analysis.*

*In particular, this proves that there exists a Hall set exceeding the general geometric  $\theta$ -based lower bound of Proposition 5.7 and the geometric  $\theta$ -based upper bounds which were valid for length-compatible and Lyndon Hall sets.*

## 10 Asymmetric estimates

The main goal of this section is to prove Theorem 1.9 which provides asymmetric estimates for the structure constants relative to any Hall set. In particular, this gives a positive answer in the case of Hall bases to the open problem raised in [2, Section 2.4.3] and allows to apply the conditional result in [2, Section 4.4.3]) to such bases, which was our main motivation.

Throughout this section,  $\mathcal{B} \subset \text{Br}(X)$  is a Hall set where we single out the role of a particular indeterminate  $X_0 \in X$ . For  $b \in \mathcal{B}$ , we denote by  $n_0(b)$  the number of occurrences of  $X_0$  in  $b$ , and  $n(b) := |b| - n_0(b)$  the number of leaves of  $b$  that are different from  $X_0$ .

We start with preliminary definitions in Section 10.1 where we introduce a way to represent trees which reflects the asymmetry between  $X_0$  and  $X \setminus \{X_0\}$ . Then, in Section 10.2, we prove estimates for the norm of such brackets, culminating with Proposition 10.9. Eventually, we explain in Section 10.3 how these notions allow to prove Theorem 1.9.

### 10.1 An asymmetric representation of trees

We start by defining a new family of brackets, which stores additional  $X_0$  factors throughout the whole tree and will allow us to manipulate these  $X_0$  factors differently from the non- $X_0$  leaves.

**Definition 10.1** (Weighted brackets). *Let  $A \subset \mathcal{B}$  with  $X_0 \notin A$ . We define  $\text{Br}^*(A)$  by induction as follows, together with maps  $\rho$  and  $\omega$  from  $\text{Br}^*(A)$  to  $\mathbb{N}$ .*

- For each  $a \in A$  and  $\nu \in \mathbb{N}$ , the pair  $\ell := a, \nu$  is called a leaf and belongs to  $\text{Br}^*(A)$ . We will use the notations  $\ell = a0^\nu$  (instead of the pair notation),  $\omega(\ell) = \rho(\ell) := \nu$ ,  $\alpha(\ell) := a$ ,  $|\ell| := 1$  and  $L(\ell) := \{a\}$ .
- For each  $t_1, t_2 \in \text{Br}^*(A)$  and  $\nu \in \mathbb{N}$ , the triple  $t := t_1, t_2, \nu$  belongs to  $\text{Br}^*(A)$ . We will use the notations  $t = \langle t_1, t_2 \rangle 0^\nu$ ,  $\omega(t) := \nu$ ,  $\rho(t) := \rho(t_1) + \rho(t_2) + \omega(t)$ ,  $|t| := |t_1| + |t_2|$  and  $L(t) := L(t_1) \cup L(t_2)$ .

**Example 10.2.** *If  $A = \{a_1, a_2, a_3\}$ ,  $t := \langle a_1 0^2, \langle a_2 0^7, a_1 \rangle 0^4 \rangle 0^3 \in \text{Br}^*(A)$  and can be visualised as*

$$\begin{array}{c}
 0^3 \\
 \swarrow \quad \searrow \\
 a_1 0^2 \quad 0^4 \\
 \quad \quad \swarrow \quad \searrow \\
 \quad \quad a_2 0^7 \quad a_1
 \end{array}
 \tag{10.1}$$

*In particular,  $|t| = 3$  and  $L(t) = \{a_1, a_2\}$ .*

**Remark 10.3.** *In this section, we use the notation  $\langle \cdot, \cdot \rangle$  to denote the weighted bracket in  $\text{Br}^*(A)$ , while, in all previous sections of this paper, this notation is used for the bracket in  $\text{Br}(\text{Br}(X))$ . No confusion is possible here since we only manipulate  $\text{Br}^*(A)$  in this section, and not  $\text{Br}(\text{Br}(X))$ . Also, we used the notations  $|t|$  and  $L(t)$  to denote the length and leaves of  $t \in \text{Br}^*(A)$ , without indexing them by  $A$ , to avoid overloading the formulas in the sequel. Again, no confusion is possible in this section.*

**Definition 10.4** (Canonical evaluation). *Let  $A \subset \mathcal{B}$  with  $X_0 \notin A$ . There is a natural evaluation mapping  $\mathfrak{r}^*$  from  $\text{Br}^*(A)$  to  $\text{Br}(X)$ , defined by induction as*

- for a leaf  $\ell = a0^\nu$  with  $a \in A$  and  $\nu \in \mathbb{N}$ ,  $\mathfrak{r}^*(\ell) := \text{ad}_{X_0}^\nu(a)$ ,
- for a tree  $t = \langle t_1, t_2 \rangle 0^\nu$  with  $t_1, t_2 \in \text{Br}^*(A)$  and  $\nu \in \mathbb{N}$ ,  $\mathfrak{r}^*(t) := \text{ad}_{X_0}^\nu((\mathfrak{r}^*(t_1), \mathfrak{r}^*(t_2)))$ .

*So elements of  $\text{Br}^*(A)$  can be evaluated in  $\text{Br}(X)$  then in  $\mathcal{L}(X)$ .*

**Definition 10.5.** *We define*

- $\mathcal{B}_\# := \mathcal{B} \setminus \{X_0\}$ ,
- $\text{ABr}^*(\mathcal{B}_\#)$  the subset of trees  $t \in \text{Br}^*(\mathcal{B}_\#)$  such that  $\{X_0\} \cup L(t)$  is alphabetic.

Along the decomposition algorithm, the additional  $X_0$  factors of the considered trees will “move upwards”. To track this phenomenon, we introduce the following definition.

**Definition 10.6** (Total depth of the  $X_0$  factors). *For  $t \in \text{Br}^*(\mathcal{B}_\#)$  and  $d \in \mathbb{N}$ , let*

$$P_d(t) := \begin{cases} d\nu & \text{if } t = a0^\nu \text{ is a leaf,} \\ d\nu + P_{d+1}(t_1) + P_{d+1}(t_2) & \text{if } t = \langle t_1, t_2 \rangle 0^\nu. \end{cases} \quad (10.2)$$

Hence,  $P(t) := P_0(t)$  represents the sum of the 0-based depths of the additional  $X_0$  factors in  $t$ . As an example, if  $t = \langle a_1 0^2, \langle a_2 0^7, a_3 \rangle 0^4 \rangle 0^3$ ,  $P(t) = 0 \times 3 + (1 \times 2) + (1 \times 4 + 2 \times 7 + 2 \times 0) = 20$ .

## 10.2 Estimates for the norm of asymmetric trees

The goal of this paragraph is to prove Proposition 10.9, which yields an estimate of the norm of (evaluations of) brackets of  $\text{ABr}^*(\mathcal{B}_\#)$ . We start with a particular case in Lemma 10.7 and a decomposition result in Lemma 10.8.

**Lemma 10.7.** *Let  $t \in \text{ABr}^*(\mathcal{B}_\#)$ . Assume that  $X_0 < \min L(t)$ . Then*

$$\|\mathbb{E}(\Gamma^*(t))\|_{\mathcal{B}} \leq |t|^{\rho(t)} (|t| - 1)!. \quad (10.3)$$

*Proof.* Since  $X_0$  is minimal, the proof consists in “pushing” all the additional  $X_0$  factors all the way down on the leaves of  $t$  by iterating the Jacobi identity. We introduce

$$G := \{\text{ad}_{X_0}^k(g); g \in L(t), k \in \mathbb{N}\} \subset \mathcal{B}_\#. \quad (10.4)$$

Since  $\{X_0\} \cup L(t)$  is alphabetic and  $X_0$  is minimal within this set, iterating Lemma 3.6 proves that  $G$  is alphabetic. Hence, by Theorem 3.10, for any  $h \in \text{Br}(G)$ ,  $\|\mathbb{I}(h)\|_{\mathcal{B}} \leq (|h| - 1)!$ . Thus, to prove (10.3), it is sufficient to prove that  $\mathbb{E}(\Gamma^*(t))$  is the sum of at most  $|t|^{\rho(t)}$  evaluations of brackets of  $\text{Br}(G)$  of length  $|t|$ .

We proceed by induction on  $|t| \geq 1$ . When  $|t| = 1$ ,  $t = g0^\nu$ , so, by definition,  $\Gamma^*(t) = \mathbb{I}(h)$  where  $h \in \text{Br}(G)$  is the leaf  $\text{ad}_{X_0}^\nu(g) \in G$ . Assume that  $|t| \geq 2$ , then  $t = \langle t_1, t_2 \rangle 0^\nu$  with  $t_1, t_2 \in \text{Br}^*(\mathcal{B}_\#)$ ,  $|t_1| + |t_2| = |t|$ ,  $\rho(t) = \rho(t_1) + \rho(t_2) + \nu$ . The Leibniz formula proves that

$$\mathbb{E}(\Gamma^*(t)) = \sum_{j=0}^{\nu} \binom{\nu}{j} [\Gamma^*(t_1 0^j), \Gamma^*(t_2 0^{\nu-j})]. \quad (10.5)$$

Since  $|t_1 0^j| < |t|$  and  $|t_2 0^{\nu-j}| < |t|$ , we can apply the induction hypothesis. Hence, we know that  $t_1 0^j$  is the sum of at most  $|t_1|^{\rho(t_1)+j}$  elements of  $\text{Br}(G)$  of length  $|t_1|$ , and  $t_2 0^{\nu-j}$  of at most  $|t_2|^{\rho(t_2)+\nu-j}$  elements of  $\text{Br}(G)$  of length  $|t_2|$ . Moreover, since

$$\sum_{j=0}^{\nu} \binom{\nu}{j} |t_1|^{\rho(t_1)+j} |t_2|^{\rho(t_2)+\nu-j} = |t_1|^{\rho(t_1)} |t_2|^{\rho(t_2)} (|t_1| + |t_2|)^\nu \leq |t|^{\rho(t)}, \quad (10.6)$$

we conclude that  $t$  is indeed the sum of at most  $|t|^{\rho(t)}$  elements of  $\text{Br}(G)$  of length  $|t|$ .  $\square$

**Lemma 10.8.** *Let  $t \in \text{ABr}^*(\mathcal{B}_\#)$  with  $|t| \geq 2$ . Assume that there exists a leaf  $\ell$  of  $t$  such that  $\alpha(\ell) = \min L(t) < X_0$  and  $\omega(\ell) = 0$ . Then*

$$\pm \mathbb{E}(\Gamma^*(t)) = \sum_{k=1}^q \mathbb{E}(\Gamma^*(h_k)) + \sum_{k=1}^r \mathbb{E}(\Gamma^*(t_k)) \quad (10.7)$$

where  $q \in \llbracket 1, |t| - 1 \rrbracket$ ,  $r \in \llbracket 0, \rho(t) \rrbracket$  and

- $h_k \in \text{Br}^*(L(t) \cup (\alpha(\ell), L(t)))$  with  $|h_k| = |t| - 1$ ,  $\rho(h_k) = \rho(t)$  and  $P(h_k) \leq P(t)$ ,
- $t_k \in \text{Br}^*(L(t))$  with  $|t_k| = |t|$ ,  $\rho(t_k) = \rho(t) - 1$  and  $P(t_k) \leq P(t) - k$ .

In particular,  $h_k \in \text{ABr}^*(\mathcal{B}_\#)$  by Lemma 3.6.

*Proof.* Let  $g := \alpha(\ell)$  be the minimal element of  $L(t)$ . Heuristically, the decomposition (10.7) comes from distributing  $g$  over all the non- $X_0$  leaves and  $X_0$  factors of its neighbor thanks to the Jacobi identity. The first sum corresponds to terms where  $g$  hits a non- $X_0$  leaf of  $t$ , yielding a shorter tree on a new alphabetic subset. The second sum corresponds to terms where  $g$  hits an additional  $X_0$  factor. An appropriate ordering of the created brackets  $t_k$  yields the delicate point of the estimate  $P(t_k) \leq P(t) - k$ , which proves that the  $X_0$  factors move upwards in this process.

*Step 1: Case when  $t = \langle g, w \rangle$ .* We proceed by induction on  $|w| \geq 1$ . We actually propagate the following stronger equality (without the sign alternative in (10.7))

$$\mathbb{E}(\mathbf{I}^*(t)) = \sum_{k=1}^q \mathbb{E}(\mathbf{I}^*(h_k)) + \sum_{k=1}^r \mathbb{E}(\mathbf{I}^*(t_k)) \quad (10.8)$$

and the stronger conditions that for all  $d \in \mathbb{N}$ ,  $P_d(h_k) \leq P_d(t)$  and  $P_d(t_k) \leq P_d(t) - k$ . We will use that, since we assumed  $\omega(\ell) = 0$ ,  $P_d(g) = 0$ .

First, if  $|w| = 1$ , then  $w = h0^\nu$  for some  $h \in \mathcal{B}_\#$  and  $\nu := \omega(w) \geq 0$ . Distributing  $g$  using the Jacobi identity yields

$$\mathbb{E}(\mathbf{I}^*(\langle g, h0^\nu \rangle)) = \mathbb{E}(\mathbf{I}^*(\langle g, h \rangle 0^\nu)) + \sum_{k=1}^{\nu} \mathbb{E}(\mathbf{I}^*(\langle \langle g, X_0 \rangle, h0^{\nu-k} \rangle 0^{k-1})), \quad (10.9)$$

which is (10.7) with  $h_1 := \langle g, h \rangle 0^\nu$  and  $t_k := \langle \langle g, X_0 \rangle, h0^{\nu-k} \rangle 0^{k-1}$ . Moreover, for every  $d \in \mathbb{N}$ ,  $P_d(\langle g, h0^\nu \rangle) = (d+1)\nu$  and  $P_d(t_k) = d(k-1) + (d+1)(\nu-k) = P_d(t) - k - d$ . So  $P_d(t_k) \leq P_d(t) - k$ .

Now assume that  $|w| \geq 2$ . Then  $w = \langle w_1, w_2 \rangle 0^\nu$  with  $w_1, w_2 \in \text{Br}^*(L(w))$  and  $\nu := \omega(w) \geq 0$ . Distributing  $g$  using the Jacobi identity yields

$$\begin{aligned} \mathbb{E}(\mathbf{I}^*(\langle g, w \rangle)) &= \mathbb{E}(\mathbf{I}^*(\langle \langle g, w_1 \rangle, w_2 \rangle 0^\nu)) + \mathbb{E}(\mathbf{I}^*(\langle w_1, \langle g, w_2 \rangle \rangle 0^\nu)) \\ &\quad + \sum_{k=1}^{\nu} \mathbb{E}(\mathbf{I}^*(\langle \langle g, X_0 \rangle, \langle w_1, w_2 \rangle 0^{\nu-k} \rangle 0^{k-1})) \end{aligned} \quad (10.10)$$

We apply the induction hypothesis to  $\langle g, w_1 \rangle$  and  $\langle g, w_2 \rangle$ , indexing all variables with 1 or 2 respectively. Thus

$$\begin{aligned} \mathbb{E}(\mathbf{I}^*(\langle g, w \rangle)) &= \sum_{k=1}^{q_1} \mathbb{E}(\mathbf{I}^*(\langle h_k^1, w_2 \rangle 0^\nu)) + \sum_{k=1}^{q_2} \mathbb{E}(\mathbf{I}^*(\langle w_1, h_k^2 \rangle 0^\nu)) \\ &\quad + \sum_{k=1}^{r_1} \mathbb{E}(\mathbf{I}^*(\langle t_k^1, w_2 \rangle 0^\nu)) + \sum_{k=1}^{r_2} \mathbb{E}(\mathbf{I}^*(\langle w_1, t_k^2 \rangle 0^\nu)) \\ &\quad + \sum_{k=1}^{\nu} \mathbb{E}(\mathbf{I}^*(\langle \langle g, X_0 \rangle, \langle w_1, w_2 \rangle 0^{\nu-k} \rangle 0^{k-1})). \end{aligned} \quad (10.11)$$

Let us check that this is the expected decomposition.

- The first two sums yield the first sum of (10.7). One checks that

$$- |\langle h_k^1, w_2 \rangle 0^\nu| = |h_k^1| + |w_2| = |\langle g, w_1 \rangle| - 1 + |w_2| = |t| - 1,$$

- $\rho(\langle h_k^1, w_2 \rangle 0^\nu) = \nu + \rho(h_k^1) + \rho(w_2) = \nu + \rho(\langle g, w_1 \rangle) + \rho(w_2) = \rho(t)$ ,
  - for every  $d \in \mathbb{N}$ ,  $P_d(\langle h_k^1, w_2 \rangle 0^\nu) = d\nu + P_{d+1}(h_k^1) + P_{d+1}(w_2) \leq d\nu + P_{d+1}(\langle g, w_1 \rangle) + P_{d+1}(w_2) = d\nu + P_{d+2}(w_1) + P_{d+1}(w_2) \leq (d+1)\nu + P_{d+1}(\langle w_1, w_2 \rangle) = P_d(t)$ ,
  - and similarly for the trees  $\langle w_1, h_k^2 \rangle 0^\nu$ ,
  - and eventually,  $q_1 + q_2 = |w_1| + |w_2| = |t| - 1$ .
- The last three sums yield the second sum of (10.7). The properties on the lengths, the leaves and the values of  $\rho$  of the trees are straightforward. We focus on the delicate estimate of the values of  $P$ . Let  $d \in \mathbb{N}$ .

- First,  $P_d(t) = P_d(\langle g, \langle w_1, w_2 \rangle 0^\nu \rangle) = (d+1)\nu + P_{d+2}(w_1) + P_{d+2}(w_2)$ .
- For  $k \in \llbracket 1, \nu \rrbracket$ ,  $P_d(\langle (g, X_0), \langle w_1, w_2 \rangle 0^{\nu-k} \rangle 0^{k-1}) = d(k-1) + (d+1)(\nu-k) + P_{d+2}(w_1) + P_{d+2}(w_2) = P_d(t) - k - d \leq P_d(t) - k$ .
- For  $k \in \llbracket 1, r_1 \rrbracket$ , since, by the induction hypothesis,  $P_{d+1}(t_k^1) \leq P_{d+1}(\langle g, w_1 \rangle) - k = P_{d+2}(w_1) - k$ , we have  $P_d(\langle t_k^1, w_2 \rangle 0^\nu) = d\nu + P_{d+1}(w_2) + P_{d+1}(t_k^1) \leq d\nu + P_{d+2}(w_2) - \rho(w_2) + P_{d+2}(w_1) - k = P_d(t) - \nu - \rho(w_2) - k$ .
- Similarly, for  $k \in \llbracket 1, r_2 \rrbracket$ ,  $P_d(\langle w_1, t_k^2 \rangle 0^\nu) \leq P_d(t) - \nu - \rho(w_1) - k$ .

Recall that  $r_1 \leq \rho(w_1)$  and  $r_2 \leq \rho(w_2)$ . Let  $r := \nu + r_1 + r_2$  and define, for  $k \in \llbracket 1, r \rrbracket$

$$t_k := \begin{cases} \langle (g, X_0), \langle w_1, w_2 \rangle 0^{\nu-k} \rangle 0^{k-1} & \text{for } k \in \llbracket 1, \nu \rrbracket, \\ \langle t_{k-\nu}^1, w_2 \rangle 0^\nu & \text{for } k \in \llbracket \nu + 1, \nu + r_1 \rrbracket, \\ \langle w_1, t_{k-\nu-r_1}^2 \rangle 0^\nu & \text{for } k \in \llbracket \nu + r_1 + 1, \nu + r_1 + r_2 \rrbracket. \end{cases} \quad (10.12)$$

The previous estimates prove that, for this labeling,  $P_d(t_k) \leq P_d(t) - k$ .

*Step 2: General case.* Up to the left-right anti-symmetry (so up to a sign in  $\mathcal{L}(X)$ ), we can assume that there is a  $w \in \text{Br}^*(L(t))$  and  $\mu \geq 0$  such that  $t'0^\mu$  where  $t' := \langle g, w \rangle$  is a subtree of  $t$ . By the previous step,

$$\mathbb{E}(\mathbb{I}^*(t')) = \sum_{k=1}^{q'} \mathbb{E}(\mathbb{I}^*(h'_k)) + \sum_{k=1}^{r'} \mathbb{E}(\mathbb{I}^*(t'_k)) \quad (10.13)$$

where  $q' \in \llbracket 1, |w| \rrbracket$ ,  $r' \in \llbracket 0, \rho(w) \rrbracket$  and

- $h'_k \in \text{Br}^*(L(w) \cup (g, L(w)))$  with  $|h'_k| = |w|$ ,  $\rho(h'_k) = \rho(w)$  and  $P(h'_k) \leq P(w)$ ,
- $t'_k \in \text{Br}^*(L(w))$  with  $|t'_k| = |w| + 1$ ,  $\rho(t'_k) = \rho(w) - 1$  and  $P(t'_k) \leq P(w) - k$ .

This proves (10.7) where  $h_k$  is obtained by replacing  $t'$  by  $h'_k$  in  $t$ , and  $t_k$  is obtained by replacing  $t'$  by  $t'_k$  in  $t$ . The properties on the leaves, the lengths and the values of  $\rho$  for the  $h_k$  and  $t_k$  can be checked easily. The only delicate part is to verify the values of  $P$ . Let  $d$  be the 0-based depth of  $t'$  in  $t$ . Then  $P(h_k) = P(t) - P_d(t') + P_d(h'_k) \leq P(t)$  because we proved  $P_d(h'_k) \leq P_d(t')$ . Similarly,  $P(t_k) = P(t) - P_d(t') + P_d(t'_k) \leq P(t) - k$ .  $\square$

**Proposition 10.9.** *Let  $t \in \text{ABr}^*(\mathcal{B}_\mathbb{q})$ . Then*

$$\|\mathbb{E}(\mathbb{I}^*(t))\|_{\mathcal{B}} \leq C(|t|)^{\rho(t)} (|t| - 1)!, \quad (10.14)$$

where, for  $n \in \mathbb{N}^*$ ,

$$C(n) := \begin{cases} 1 & \text{for } n = 1, \\ 2^{\frac{n(n-1)}{2} + 1} & \text{when } n \geq 2. \end{cases} \quad (10.15)$$

*Proof.* For  $n \geq 1$ ,  $\rho \geq 0$  and  $P \geq 0$ , we introduce

$$F_n(\rho, P) := \max \left\{ \frac{\|\mathbb{E}(\Gamma^*(t))\|_{\mathcal{B}}}{(|t|-1)!}; t \in \text{ABr}^*(\mathcal{B}_{\sharp}), |t| = n, \rho(t) \leq \rho, P(t) \leq P \right\}. \quad (10.16)$$

In particular, for each  $n \geq 1$ ,  $F_n$  is a non-decreasing function of  $\rho$  and  $P$ . Moreover, since, for every  $t \in \text{Br}^*(\mathcal{B}_{\sharp})$ ,  $P(t) \leq (|t|-1)\rho(t)$ , we have, for every  $P \geq 0$ ,

$$F_n(\rho, P) \leq \bar{F}_n(\rho) := F_n(\rho, (n-1)\rho). \quad (10.17)$$

*Step 1: We compute the values of  $\bar{F}_n(\rho)$  for  $n = 1$  or  $\rho = 0$ .* First, for every  $\rho \geq 0$ ,

$$\bar{F}_1(\rho) = 1. \quad (10.18)$$

The proof is by induction on  $\rho \geq 0$ . Let  $t \in \text{ABr}^*(\mathcal{B}_{\sharp})$  such that  $|t| = 1$  and  $\rho(t) = \rho$ , so  $t = g0^{\rho}$  for some  $g \in \mathcal{B}_{\sharp}$ . If  $\rho = 0$ ,  $t$  is already a leaf labeled by  $g \in \mathcal{B}$ , without additional  $X_0$  factors, so  $\Gamma^*(t) = g \in \mathcal{B}$  and thus  $\|\mathbb{E}(\Gamma^*(t))\|_{\mathcal{B}} = 1$ . Otherwise,  $\rho > 0$ . If  $X_0$  is minimal among  $\{X_0, g\}$ ,  $\Gamma^*(t) = \text{ad}_{X_0}^{\rho}(g) \in \mathcal{B}$  so  $\|\mathbb{E}(\Gamma^*(t))\|_{\mathcal{B}} = 1$ . If  $g$  is minimal among  $\{X_0, g\}$ ,  $\mathbb{E}(\Gamma^*(t)) = -\mathbb{E}(\Gamma^*((g, X_0)0^{\rho-1}))$ , so the proof follows by induction.

Second, for every  $n \geq 1$ ,

$$\bar{F}_n(0) \leq 1. \quad (10.19)$$

Indeed, for  $t \in \text{ABr}^*(\mathcal{B}_{\sharp})$  with  $\rho(t) = 0$  and  $|t| = n$ , since there is no additional  $X_0$  factors,  $\Gamma^*(t)$  is equal, in  $\text{Br}(X)$ , to the evaluation of a bracket of length  $n$  over an alphabetic subset of  $\mathcal{B}_{\sharp}$ . So Theorem 3.10 yields  $\|\mathbb{E}(\Gamma^*(t))\|_{\mathcal{B}} \leq (n-1)!$ .

*Step 2: Proof of a functional inequality using the Jacobi identity.* We prove that, for every  $n \geq 2$ ,  $\rho \geq 1$  and  $P \geq 0$ ,

$$F_n(\rho, P) \leq \max \left\{ n^{\rho}, F_{n-1}(\rho, P) + \sum_{k=1}^{+\infty} F_n(\rho-1, P-k) \right\}, \quad (10.20)$$

with the convention that  $F_n(\rho', P') = 0$  when  $P' < 0$ .

Let  $t \in \text{ABr}^*(\mathcal{B}_{\sharp})$  with  $|t| = n$ ,  $\rho(t) \leq \rho$  and  $P(t) \leq P$ . We separate three cases. Let  $g := \min(\{X_0\} \cup L(t))$ .

- *Case  $g = X_0$ .* Then Lemma 10.7 proves that  $\frac{\|\mathbb{E}(\Gamma^*(t))\|_{\mathcal{B}}}{(|t|-1)!} \leq n^{\rho}$ . This case yields the first term of the right-hand side maximum in (10.20).
- *Case  $g \in L(t)$  and there is a leaf  $\ell$  of  $t$  such that  $g = \alpha(\ell)$  and  $\omega(\ell) = 0$ .* Using the decomposition (10.7) with the notations of Lemma 10.8 and the monotony of  $F_{n-1}$  and  $F_n$ , we obtain

$$\begin{aligned} \frac{\|\mathbb{E}(\Gamma^*(t))\|_{\mathcal{B}}}{(n-1)!} &\leq \frac{1}{(n-1)!} \sum_{k=1}^q \|\Gamma^*(h_k)\|_{\mathcal{B}} + \frac{1}{(n-1)!} \sum_{k=1}^r \|\Gamma^*(t_k)\|_{\mathcal{B}} \\ &\leq \frac{(n-1)}{(n-1)!} (n-2)! F_{n-1}(\rho, P) + \sum_{k=1}^r F_n(\rho-1, P-k), \end{aligned} \quad (10.21)$$

which yields the second term of the right-hand side maximum in (10.20).

- *Case  $g \in L(t)$  but for each leaf  $\ell$  of  $t$  such that  $g = \alpha(\ell)$ ,  $\omega(\ell) > 0$ .* Take such a leaf and let  $\nu := \omega(\ell) > 0$ , we have  $\mathbb{E}(\Gamma^*(t)) = -\mathbb{E}(\Gamma^*(\bar{t}))$  where  $\ell = g0^{\nu}$  in  $t$  has been replaced by  $(g, X_0)0^{\nu-1}$  in  $\bar{t}$ . By Lemma 3.6,  $\{X_0\} \cup L(\bar{t})$  is alphabetic. Moreover,  $|\bar{t}| = |t|$ ,  $\rho(\bar{t}) = \rho(t) - 1$  and  $P(\bar{t}) \leq P(t)$ . Iterating this procedure if necessary brings us eventually back to the first two cases.



*Step 3: Resolution of the functional inequality and conclusion.* Let  $n \geq 2$ ,  $\rho \geq 1$  and  $P \geq 0$ . Let  $f_{n,\rho,P} := \max\{n^\rho, F_{n-1}(\rho, P)\} \in \mathbb{R}_+$ . Iterating (10.20) and using the monotony of  $f_{n,\rho,P}$  with respect to  $\rho$  and  $P$  yields

$$\begin{aligned}
F_n(\rho, P) &\leq f_{n,\rho,P} + \sum_{k_1=1}^{+\infty} F_n(\rho-1, P-k_1) \\
&\leq f_{n,\rho,P} + \sum_{k_1=1}^{+\infty} \left( f_{n,\rho-1,P-k_1} + \sum_{k_2=1}^{+\infty} F_n(\rho-2, P-k_1-k_2) \right) \\
&\leq \dots \\
&\leq f_{n,\rho,P} \left( 1 + \sum_{r=1}^{\rho-1} |\{(k_1, \dots, k_r) \in (\mathbb{N}^*)^r; k_1 + \dots + k_r \leq P\}| \right) \\
&\quad + \overline{F}_n(0) |\{(k_1, \dots, k_\rho) \in (\mathbb{N}^*)^\rho; k_1 + \dots + k_\rho \leq P\}|.
\end{aligned} \tag{10.22}$$

Using the classical bound for the number of compositions of an integer therefore yields

$$F_n(\rho, P) \leq 2^P \max\{n^\rho, F_{n-1}(\rho, P), \overline{F}_n(0)\}. \tag{10.23}$$

By (10.19),  $\overline{F}_n(0) \leq 1 \leq n^\rho$ . Thus, using (10.17),

$$\overline{F}_n(\rho) \leq 2^{(n-1)\rho} \max\{n^\rho, \overline{F}_{n-1}(\rho)\}. \tag{10.24}$$

Recalling (10.18), we obtain  $\overline{F}_2(\rho) \leq 2^\rho \max\{2^\rho, 1\} = 2^{2\rho} = C(2)^\rho$ . Then for  $n \geq 3$ , with  $C$  defined as in (10.15), one can ignore the term  $n^\rho$  in the maximum since  $n \leq C(n-1)$  for  $n \geq 3$ . We thus obtain

$$\overline{F}_n(\rho) \leq C(n)^\rho, \tag{10.25}$$

Hence, if  $t \in \text{ABr}^*(\mathcal{B}_\#)$ , recalling (10.16),

$$\|E(\Gamma^*(t))\|_{\mathcal{B}} \leq (|t| - 1)! \times \overline{F}_{|t|}(\rho(t)), \tag{10.26}$$

which, together with (10.25), concludes the proof of (10.14) for  $n \geq 2$  and  $\rho \geq 1$  (the cases  $n = 1$  or  $\rho = 0$  were already covered in the initializations step).  $\square$

### 10.3 Proof of the main generic asymmetric estimate

We prove the following refined version of Theorem 1.9. For  $a < b \in \mathcal{B}$ , let  $\rho_a(b)$  denote the number of occurrences of  $X_0$  as a leaf of  $T_a(b)$  and  $n_a(b) := \theta_a(b) - \rho_a(b)$ . In particular, by construction,  $\rho_a(b) \leq n_0(b)$  and  $n_a(b) \leq n(b)$ .

**Theorem 10.10.** *Let  $a < b \in \mathcal{B}$ . Then*

$$\|[a, b]\|_{\mathcal{B}} \leq C(n_a(b) + 1)^{\rho_a(b)} n_a(b)!, \tag{10.27}$$

where  $C$  is the non-decreasing sequence defined in (10.15).

*Proof.* Let  $a < b \in \mathcal{B}$ . When  $b = X_0$ , by the axioms of a Hall set,  $(a, b) \in \mathcal{B}$  so  $\|[a, b]\|_{\mathcal{B}} = 1$ , which corresponds to the estimate with  $n_a(b) = 0$ ,  $\rho_a(b) = 1$  and  $C(1) = 1$ . When  $b \neq X_0$ , we separate two cases.

- *Case  $X_0 \notin L_{\mathcal{B}}(T_a(b))$*  (which happens in particular when  $a = X_0$  or if  $X_0$  is minimal in  $\mathcal{B}$ ). Then  $\rho_a(b) = 0$  and  $\theta_a(b) = n_a(b)$ . By Proposition 4.3,  $\{a\} \cup L_{\mathcal{B}}(T_a(b))$  is an alphabetic subset of  $\mathcal{B}$ . Seeing  $(a, b)$  as  $\mathbf{1}(\langle a, T_a(b) \rangle)$ , a bracket of  $\theta_a(b) + 1 = n_a(b) + 1$  elements of an alphabetic subset, yields  $\|[a, b]\|_{\mathcal{B}} \leq n_a(b)!$  by Theorem 3.10.

- *Case*  $X_0 \in L_{\mathcal{B}}(T_a(b))$  (which implies  $a \neq X_0$ ). Let  $A := (\{a\} \cup L_{\mathcal{B}}(T_a(b))) \setminus \{X_0\} \subset \mathcal{B}_{\sharp}$ . Then, collecting all the  $X_0$  factors in the map  $\omega$ , there exists a tree  $t \in \text{Br}^*(A)$  with  $|t| = n_a(b) + 1$  and  $\rho(t) = \rho_a(b)$  such that  $\text{E}(\mathbf{1}(\langle a, T_a(b) \rangle)) = \pm \text{E}(\mathbf{1}^*(t))$ . Moreover,  $L(t) \cup \{X_0\} = A \cup \{X_0\} = \{a\} \cup L_{\mathcal{B}}(T_a(b))$  is alphabetic by Proposition 4.3. So  $t \in \text{ABr}^*(\mathcal{B}_{\sharp})$  and the estimate follows from Proposition 10.9.

Hence, (10.27) is valid in both cases.  $\square$

**Remark 10.11.** *Estimate (10.27) implies that, for each Hall set and with respect to each indeterminate, the asymmetric asymptotic growth is at most geometric.*

*Theorem 5.5 proves that, for each Hall set, the asymmetric asymptotic growth with respect to  $\max X$  is at least geometric.*

*Since, in each Hall set,  $\min X = \min \mathcal{B}$ , the first case of the proof of Theorem 10.10 implies that the asymmetric asymptotic growth with respect to  $\min X$  is in fact bounded.*

## 11 Perspectives and open problems

We present some possible extensions related with the growth of structure constants of free Lie algebras, which we find interesting.

In Section 2.1, we mentioned that structure constants are related with the (space and time) complexity of the rewriting algorithm which allows to decompose  $[a, b]$  on the basis. In this direction, one could start by considering that the comparison operation  $a < b$  and the bracket creation operation are elementary operations of time-complexity  $\mathcal{O}(1)$  and then investigate the time-complexity of the `Rewrite` function described in Proposition 4.17.

In Section 4, we proved that  $\|[a, b]\|_{\mathcal{B}} \leq \lfloor e(n-1)! \rfloor$  when  $|b| = n$  and that this estimate is sharp in the sense of Corollary 4.15. The asymptotic optimality case for large  $n$  uses an infinite set  $X$ . When  $|X| = 2$ , our construction of Section 9 only provides an asymptotic growth behaving roughly like  $\sqrt{n}^n$  (see (9.14)), which is quite far from  $n^n$ . Therefore, an interesting direction would be to investigate the dependency of the optimal estimates with respect to the cardinal of  $X$ .

The examples in Section 5 prove that the structure constants of free Lie algebras relative to Hall bases grow at least geometrically with the length of the involved brackets. Hence, in this sense, the product operation has an exponential size within these bases. An interesting open problem would be to determine if there exist other bases of  $\mathcal{L}(X)$  (see Appendix E for a short discussion) having a bounded size, a polynomial size, or at least a sub-exponential size, in the sense that

$$\limsup_{n \rightarrow +\infty} \|[a, b]\|_{\mathcal{B}} < +\infty, \quad \limsup_{n \rightarrow +\infty} \frac{\ln \|[a, b]\|_{\mathcal{B}}}{\ln n} < +\infty \quad \text{or} \quad \limsup_{n \rightarrow +\infty} \frac{\ln \|[a, b]\|_{\mathcal{B}}}{n} = 0, \quad (11.1)$$

where the suprema are taken over  $a$  and  $b$  elements of the basis such that  $[a, b] \in \mathcal{L}_n(X)$ .

In Section 10, we proved asymmetric estimates of the form  $C(n_a(b) + 1)^{\rho_a(b)} n_a(b)!$ , where  $C(n)$  behaves roughly like  $2^{n^2}$  (see (10.15)). It would be interesting to investigate whether this behavior is optimal or if one can improve our proof to obtain a better dependency on  $n$  of the geometric rate  $C(n)$ , along with a “critical asymmetric basis” which achieves this rate. One could also investigate how the maximal or minimal growth of  $C(n)$  depends on  $|X|$ .

Eventually, this paper focuses mainly on the “worst case” size of the Lie bracket operation within Hall bases. From the point of view of computational applications, it could also be interesting to investigate its average size for fixed length, e.g. through the quantities

$$\gamma_n(\mathcal{B}) := \text{mean} \{ \|[a, b]\|_{\mathcal{B}}; a < b \in \mathcal{B}, |a| + |b| = n \}. \quad (11.2)$$

What is the asymptotic growth of  $\gamma_n(\mathcal{B})$ ? As in our case, one could start by considering this question for the classical length-compatible or Lyndon bases.

## A Detailed example of the decomposition algorithm

In this paragraph, to facilitate the comprehension of the proof of Theorem 4.12, we propose an illustration of the execution of the construction of  $B_\pi$  along the path  $\pi = (4, 5, 7)$  for  $T_a(b)$  having the structure of (4.8). The construction starts with  $B_\emptyset = \langle a, T_a(b) \rangle$ , so

$$B_\emptyset = \begin{array}{c} \diagup \quad \diagdown \\ a \end{array} \begin{array}{c} \diagup \quad \diagdown \\ \diagup \quad \diagdown \\ b_3 \quad b_1 \quad b_2 \quad b_4 \end{array} \begin{array}{c} \diagup \quad \diagdown \\ \diagup \quad \diagdown \\ b_5 \quad b_6 \quad b_7 \end{array} \quad b_8 \quad (A.1)$$

Then,  $a$  is pushed on  $b_4$ . So

$$B_{(4)} = \begin{array}{c} \diagup \quad \diagdown \\ b_3 \end{array} \begin{array}{c} \diagup \quad \diagdown \\ b_1 \quad b_2 \end{array} \begin{array}{c} \diagup \quad \diagdown \\ (a, b_4) \end{array} \begin{array}{c} \diagup \quad \diagdown \\ b_5 \quad b_6 \quad b_7 \end{array} \quad b_8 \quad (A.2)$$

Then, assuming that it is minimal,  $(a, b_4)$  is pushed on  $b_5$ . So

$$B_{(4,5)} = \begin{array}{c} \diagup \quad \diagdown \\ b_3 \end{array} \begin{array}{c} \diagup \quad \diagdown \\ b_1 \quad b_2 \end{array} \begin{array}{c} \diagup \quad \diagdown \\ ((a, b_4), b_5) \end{array} \begin{array}{c} \diagup \quad \diagdown \\ b_6 \quad b_7 \end{array} \quad b_8 \quad (A.3)$$

Eventually, assuming that  $((a, b_4), b_5)$  is minimal, it is pushed on  $b_7$ . So

$$B_{(4,5,7)} = \begin{array}{c} \diagup \quad \diagdown \\ b_3 \end{array} \begin{array}{c} \diagup \quad \diagdown \\ b_1 \quad b_2 \end{array} \begin{array}{c} \diagup \quad \diagdown \\ b_6 \end{array} \begin{array}{c} \diagup \quad \diagdown \\ a_{4,5,7} \end{array} \quad b_8 \quad (A.4)$$

where  $a_{(4,5,7)} = (((a, b_4), b_5), b_7)$ . The process stops here since the sibling of  $a_{4,5,7}$  in  $B_{(4,5,7)}$  is a leaf, so there is no longer path starting with  $(4, 5, 7)$ .

## B A Leibniz-type inversion rule

We prove a formula linked with the general Leibniz rule, which is used throughout the paper to balance iterated Lie brackets away from a particular term.

**Definition B.1** (Multinomial coefficient). For  $k \in \mathbb{N}^*$ ,  $j_1, \dots, j_k \in \mathbb{N}$  and  $\nu = j_1 + \dots + j_k$ , we define the multinomial coefficient

$$\binom{\nu}{j_1, \dots, j_k} := \frac{\nu!}{j_1! \cdots j_k!}. \quad (B.1)$$

Moreover, for all  $\nu \geq 0$ ,

$$\sum_{j_1 + \dots + j_k = \nu} \binom{\nu}{j_1, \dots, j_k} = k^\nu \quad (B.2)$$

**Lemma B.2.** *Let  $\mathcal{A}$  be an algebra and  $D$  a derivation on  $\mathcal{A}$ . For every  $\nu \in \mathbb{N}$  and  $b_1, b_2 \in \mathcal{A}$ ,*

$$b_1(D^\nu b_2) = \sum_{j=0}^{\nu} (-1)^j \binom{\nu}{j} D^{\nu-j}((D^j b_1)b_2). \quad (\text{B.3})$$

*Proof.* The proof is by induction on  $\nu \geq 0$ . The case  $\nu = 0$  is trivial. We assume the property up to some  $\nu \in \mathbb{N}$  and prove it for  $\nu + 1$ . Since  $D$  is a derivation,

$$b_1(D^{\nu+1}b_2) = D(b_1(D^\nu b_2)) - (Db_1)(D^\nu b_2). \quad (\text{B.4})$$

Applying the induction hypothesis on both terms and Pascal's formula concludes the proof.  $\square$

**Lemma B.3.** *Let  $\mathcal{A}$  be an algebra and  $D$  a derivation on  $\mathcal{A}$ . For  $k \geq 2$ ,  $\nu \in \mathbb{N}$  and  $b_1, \dots, b_k \in \mathcal{A}$ ,*

$$b_1 \cdots b_{k-1}(D^\nu b_k) = \sum_{j_1 + \cdots + j_k = \nu} (-1)^{\nu - j_k} \binom{\nu}{j_1, \dots, j_k} D^{j_k}((D^{j_1} b_1) \cdots (D^{j_{k-1}} b_{k-1})b_k), \quad (\text{B.5})$$

where all products of  $k$  terms are understood as being right-nested if  $\mathcal{A}$  is non-associative (i.e. by convention in this result  $abcd$  denotes  $(a(b(cd)))$ ).

*Proof.* The case  $k = 2$  is covered in Lemma B.2. We now proceed by induction on  $k$ , assuming that the formula holds for some  $k \geq 2$ , and proving it for  $k + 1$  elements. By the induction hypothesis and linearity, one has:

$$b_1 b_2 \cdots b_k (D^\nu b_{k+1}) = \sum_{j_2 + \cdots + j_{k+1} = \nu} (-1)^{\nu - j_{k+1}} \binom{\nu}{j_2, \dots, j_{k+1}} b_1 D^{j_{k+1}}((D^{j_2} b_2) \cdots (D^{j_k} b_k) b_{k+1}) \quad (\text{B.6})$$

Applying the two-elements case to each term yields

$$\begin{aligned} & b_1 D^{j_{k+1}}((D^{j_2} b_2) \cdots (D^{j_k} b_k) b_{k+1}) \\ &= \sum_{j_1 + j'_{k+1} = j_{k+1}} (-1)^{j_1} \binom{j_{k+1}}{j_1} D^{j'_{k+1}}((D^{j_1} b_1)(D^{j_2} b_2) \cdots (D^{j_k} b_k) b_{k+1}). \end{aligned} \quad (\text{B.7})$$

If  $j_2 + \cdots + j_{k+1} = \nu$  and  $j_1 + j'_{k+1} = j_{k+1}$ , then  $\nu = j_1 + \cdots + j_k + j'_{k+1}$ , and:

$$(-1)^{\nu - j_{k+1}} \binom{\nu}{j_2, \dots, j_{k+1}} (-1)^{j_1} \binom{j_{k+1}}{j_1} = (-1)^{\nu - j'_{k+1}} \binom{\nu}{j_1, j_2, \dots, j'_{k+1}}, \quad (\text{B.8})$$

which concludes the proof.  $\square$

## C Computations and estimates on some binomial sums

In this purely numerical section, we state and prove numerical formulas and estimates on the quantities  $A_s^r(n)$  defined in (5.15) which are involved in the derivation of our lower bounds for the growth of the structure constants. They follow from elementary manipulations of sums using famous binomial identities.

**Lemma C.1.** *Let  $1 \leq r \leq s$  and  $n \geq 2s + 1$ . Then*

$$\sum_{p=0}^{r-1} (-1)^p \binom{n-p-1}{n-s-1} \binom{n-s}{p} = (-1)^r \frac{r(r-n)}{s(n-s)} \binom{n-r-1}{n-s-1} \binom{n-s}{r}. \quad (\text{C.1})$$

*Proof.* We proceed by induction on  $r$ . For  $r = 1$ , using the absorption and symmetry identities, one checks that both sides are equal to  $\binom{n-1}{s}$ . We assume that the result holds for some  $r \geq 1$  and we prove it for  $r + 1$ . Let  $s \geq r + 1$  and  $n \geq 2s + 1$  (so in particular  $s \geq r$  so that we can apply the induction equality). Hence

$$S_{r+1} := \sum_{p=0}^{(r+1)-1} (-1)^p \binom{n-p-1}{n-s-1} \binom{n-s}{p} = (-1)^r \frac{r(r-n)}{s(n-s)} \binom{n-r-1}{n-s-1} \binom{n-s}{r} + (-1)^r \binom{n-r-1}{n-s-1} \binom{n-s}{r}. \quad (\text{C.2})$$

Using the absorption identities twice yields

$$S_{r+1} = (-1)^r \left( \frac{r(r-n)}{s(n-s)} + 1 \right) \frac{(r+1)}{n-s-r} \frac{n-(r+1)}{s-r} \binom{n-(r+1)-1}{n-s-1} \binom{n-s}{r+1}. \quad (\text{C.3})$$

Simplifying the fraction yields indeed

$$S_{r+1} = (-1)^{r+1} \frac{(r+1)((r+1)-n)}{s(n-s)} \binom{n-(r+1)-1}{n-s-1} \binom{n-s}{r+1}, \quad (\text{C.4})$$

which concludes the proof.  $\square$

**Lemma C.2.** *Let  $1 \leq r \leq s$  and  $n \geq 2s + 1$ . Then*

$$A_s^r(n) = \frac{n-2s}{s} \binom{n-r}{n-s} \binom{n-s-1}{r-1} \quad (\text{C.5})$$

*Proof.* We start with the following elementary formula, valid for  $0 \leq a \leq b$  (including the cases  $0 = a < b$  and  $a = b = 0$  with the convention that  $\binom{\cdot}{-1} = 0$ ),

$$\binom{b}{a} - \binom{b}{a-1} = \frac{b+1-2a}{b+1} \binom{b+1}{a} \quad (\text{C.6})$$

Starting from the definition (5.15) of  $A_s^r(n)$  and using (C.6) to simplify the difference and the symmetry identity for the first binomial, we obtain

$$A_s^r(n) = (n-2s) \left| \sum_{p=r}^s \frac{(-1)^p}{n-2p} \binom{n-1-p}{n-1-2p} \binom{n-2p}{s-p} \right|. \quad (\text{C.7})$$

Then, using the absorption identity for the upper index of the second binomial,

$$A_s^r(n) = (n-2s) \left| \sum_{p=r}^s \frac{(-1)^p}{n-p-s} \binom{n-1-p}{n-1-2p} \binom{n-1-2p}{s-p} \right|. \quad (\text{C.8})$$

Using the ‘‘trinomial revision’’ formula (see e.g. [19, (5.21)]) yields

$$A_s^r(n) = (n-2s) \left| \sum_{p=r}^s \frac{(-1)^p}{n-p-s} \binom{n-1-p}{s-p} \binom{n-1-s}{n-1-p-s} \right|. \quad (\text{C.9})$$

Using the absorption identity for the upper index of the second binomial and the symmetry identity twice yields

$$A_s^r(n) = \frac{n-2s}{n-s} \left| \sum_{p=r}^s (-1)^p \binom{n-1-p}{n-1-s} \binom{n-s}{p} \right|. \quad (\text{C.10})$$

Moreover, thanks to [19, (5.25)],

$$\sum_{p=0}^{n-1} (-1)^p \binom{n-1-p}{n-1-s} \binom{n-s}{p} = (-1)^s \binom{2-n}{s} = 0. \quad (\text{C.11})$$

when  $n \geq 2$ . Since the summand vanishes for  $p > s$ , we have

$$A_s^r(n) = \frac{n-2s}{n-s} \left| \sum_{p=0}^{r-1} (-1)^p \binom{n-1-p}{n-1-s} \binom{n-s}{p} \right|. \quad (\text{C.12})$$

Thus, using Lemma C.1 to compute the sum and the absorption identity twice to absorb the leading factors leads to the formula (C.5), which concludes the proof.  $\square$

**Lemma C.3.** *Let  $1 \leq r \leq s$  and  $n \geq 2s + 1$ . Then*

$$A_s^r(n) \geq \binom{n-s-1}{s}. \quad (\text{C.13})$$

*Proof.* Expanding the binomials in (C.5), inequality (C.13) is equivalent to

$$\begin{aligned} & \frac{n-2s}{s} \frac{(n-r)!}{(n-s)!(s-r)!} \frac{(n-s-1)!}{(r-1)!(n-s-r)!} \geq \frac{(n-s-1)!}{s!(n-2s-1)!} \\ \Leftrightarrow & \frac{(s-1)!(n-2s)!(n-r)!}{(n-s)!(s-r)!(r-1)!(n-s-r)!} \geq 1 \\ \Leftrightarrow & \binom{s-1}{r-1} \frac{(n-2s)!}{(n-s-r)!} \frac{(n-r)!}{(n-s)!} \geq 1 \\ \Leftrightarrow & \binom{s-1}{r-1} \prod_{j=r}^{s-1} \frac{n-j}{n-s-j} \geq 1, \end{aligned} \quad (\text{C.14})$$

which is always true since the binomial is an integer and  $n-j \geq n-s-j$  for each  $j \in \llbracket r, s-1 \rrbracket$ .  $\square$

**Lemma C.4.** *For every  $r \geq 1$ , there exists  $C_r > 0$  such that, for every  $n \geq 2r + 1$ ,*

$$A_{\lfloor \frac{n-1}{2} \rfloor}^r(n) \geq C_r n^{r-\frac{5}{2}} 2^n. \quad (\text{C.15})$$

*Proof.* Let  $r \geq 1$ . For  $n \geq 2r + 1$ , we use the notation  $m := \lfloor \frac{n-1}{2} \rfloor$ . Thus,  $n = 2m + q$  with  $q = 1$  or  $q = 2$  and  $m \geq r$ . Substituting these values in (C.5) yields

$$A_m^r(2m+q) = \frac{q}{m} \binom{2m+q-r}{m+q} \binom{m+q-1}{r-1}. \quad (\text{C.16})$$

Since  $r$  and  $q$  are fixed, Stirling's formula provides the following asymptotic for large  $m$  (or, equivalently, large  $n$ )

$$A_m^r(2m+q) \sim \frac{q}{m} \frac{(2m)^{2m+q-r} \sqrt{2\pi m}}{m^{m+q} m^{m-r} 2\pi m} \frac{m^{m+q-1} \sqrt{2\pi m}}{(r-1)! m^{m+q-r} \sqrt{2\pi m}} = \frac{qm^{r-\frac{5}{2}} 2^{2m+q-r}}{(r-1)! \sqrt{2\pi}}. \quad (\text{C.17})$$

Hence,

$$A_{\lfloor \frac{n-1}{2} \rfloor}^r(n) \sim \frac{4^{1-r} q}{(r-1)! \sqrt{\pi}} n^{r-\frac{5}{2}} 2^n, \quad (\text{C.18})$$

where  $q$  alternates between 1 and 2 so is bounded below. Since  $A_{\lfloor \frac{n-1}{2} \rfloor}^r(n) > 0$  for all values of  $n$ , choosing  $C_r$  sufficiently small concludes the proof.  $\square$

**Lemma C.5.** For every  $n \geq 3$ ,

$$1 + \sum_{s=1}^{\lfloor \frac{n-1}{2} \rfloor} A_s^1(n) = \binom{n-1}{\lfloor \frac{n-1}{2} \rfloor} \leq 2^{n-2}. \quad (\text{C.19})$$

*Proof.* Let  $n \geq 3$  and  $m := \lfloor \frac{n-1}{2} \rfloor \geq 1$ . Using (C.5) and the absorption identity,

$$\begin{aligned} \alpha_n &:= 1 + \sum_{s=1}^{\lfloor \frac{n-1}{2} \rfloor} A_s^1(n) = 1 + \sum_{s=1}^m \frac{n-2s}{s} \binom{n-1}{n-s} \\ &= 1 + \sum_{s=1}^m \frac{n}{s} \binom{n-1}{s-1} - 2 \sum_{s=1}^m \binom{n-1}{s-1} \\ &= \sum_{s=0}^m \binom{n}{s} - 2 \sum_{s=0}^{m-1} \binom{n-1}{s}. \end{aligned} \quad (\text{C.20})$$

Using elementary half summation formulas of binomial coefficients yields, when  $n = 2m + 1$ ,

$$\alpha_{2m+1} = \sum_{s=0}^m \binom{2m+1}{s} - 2 \sum_{s=0}^{m-1} \binom{2m}{s} = \frac{1}{2} 2^{2m+1} - 2 \frac{1}{2} \left( 2^{2m} - \binom{2m}{m} \right) = \binom{2m}{m}, \quad (\text{C.21})$$

and, when  $n = 2m + 2$ , thanks to the absorption identity,

$$\begin{aligned} \alpha_{2m+2} &= \sum_{s=0}^m \binom{2m+2}{s} - 2 \sum_{s=0}^{m-1} \binom{2m+1}{s} \\ &= \frac{1}{2} \left( 2^{2m+2} - \binom{2m+2}{m+1} \right) - 2 \left( \frac{1}{2} 2^{2m+1} - \binom{2m+1}{m} \right) \\ &= 2 \binom{2m+1}{m} - \frac{1}{2} \binom{2m+2}{m+1} = \binom{2m+1}{m}. \end{aligned} \quad (\text{C.22})$$

Hence, in both cases,  $\alpha_n = \binom{n-1}{m}$ , which is the claimed equality. We now prove the upper bound. Using the bounds associated with Stirling's approximation (see e.g. [34]),

$$\alpha_n = \binom{n-1}{m} \leq \frac{e^{\frac{1}{12}} \sqrt{2\pi(n-1)} (n-1)^{n-1}}{\sqrt{2\pi m m^m} \sqrt{2\pi(n-1-m)} (n-1-m)^{(n-1-m)}}. \quad (\text{C.23})$$

When  $n = 2m + 1$ , this yields

$$\alpha_{2m+1} \leq \frac{e^{\frac{1}{12}}}{\sqrt{\pi m}} 2^{2m}. \quad (\text{C.24})$$

When  $n = 2m + 2$ , this yields

$$\alpha_{2m+2} \leq \frac{e^{\frac{1}{12}}}{\sqrt{\pi m}} 2^{2m+1} \left( 1 + \frac{1}{m} \right)^m \leq \frac{e^{\frac{13}{12}}}{\sqrt{\pi m}} 2^{2m+1}. \quad (\text{C.25})$$

Hence, in both cases,  $\alpha_n \leq \frac{1}{2} 2^{n-1} = 2^{n-2}$  for  $m \geq 12$  (so  $n \geq 25$ ). One concludes the proof by checking by hand or with computer algebra that  $\alpha_n \leq 2^{n-2}$  for  $n \in \llbracket 3, 24 \rrbracket$ .  $\square$

## D Some estimates with Fibonacci numbers

As in Section 8, we denote by  $(F_\nu)_{\nu \in \mathbb{N}}$  the 0-based Fibonacci numbers. We gather here elementary estimates involving these numbers which are used in Section 8.

**Lemma D.1.** For all  $p, q \in \mathbb{N}^*$ ,

$$2F_p \leq F_{p+2}, \quad (\text{D.1})$$

$$F_p F_{q-1} + F_{p-1} F_q \leq F_{p+q-1}. \quad (\text{D.2})$$

*Proof.* Inequality (D.1) is consequence of the monotonicity of the sequence:  $F_{p+2} = F_{p+1} + F_p \geq 2F_p$ . Inequality (D.2) clearly holds if either  $p$  or  $q$  equals 1, and otherwise follows from the fact that:

$$0 \leq F_{p-2} F_{q-2} = (F_p - F_{p-1})(F_q - F_{q-1}) = F_p F_q + F_{p-1} F_{q-1} - (F_p F_{q-1} + F_{p-1} F_q) \quad (\text{D.3})$$

and from the classical identity  $F_p F_q + F_{p-1} F_{q-1} = F_{p+q-1}$  (see [43, eq. (1)]).  $\square$

**Lemma D.2.** The following estimates holds:

$$\forall n \geq 2, \quad (n-2)2^{n-2} + n \leq F_{2n-1}, \quad (\text{D.4})$$

$$\forall n \geq 2, \quad (n-3)2^{n-2} + n + 0 \leq F_{2n-2}, \quad (\text{D.5})$$

$$\forall n \geq 9, \quad (n-2)2^{n-2} + n \leq F_{2n-2} \quad (\text{D.6})$$

*Proof.* These estimates hold for large values of  $n$  because the geometric growth rate of the right-hand side is  $\varphi^2 \approx 2.618 > 2$  and can be checked numerically for small values of  $n$ .  $\square$

## E Other bases of free Lie algebras

Hall bases are of course not the only bases of  $\mathcal{L}(X)$ . For example, one can construct bases whose elements are linear combinations of Lie monomials (see e.g. [3]). For such ‘‘polynomial bases’’ one could study the growth of the structure constants with respect to  $|a| + |b|$  where  $|\cdot|$  would denote the degree of the Lie polynomial, i.e. the length of the longest Lie monomial involved.

Staying within the scope of ‘‘monomial bases’’ (i.e. bases of the form  $\mathbb{E}(\mathcal{A})$  for some  $\mathcal{A} \subset \text{Br}(X)$ , not necessarily a Hall set), one can use other construction processes than the one yielding Hall sets, as illustrated by [12] or [41]. However, even if one uses an alternative construction to obtain a monomial basis of  $\mathcal{L}(X)$ , one could wonder if there exists a Hall set yielding, up to sign, the same basis. We give below a short argument showing that this is not the case in general. Hence, there indeed exist monomial bases which cannot be seen as Hall bases.

**Proposition E.1.** Let  $X$  be a set with  $|X| \geq 2$ . There exists  $\mathcal{A} \subset \text{Br}(X)$  such that  $\mathbb{E}(\mathcal{A})$  is a basis of  $\mathcal{L}(X)$  but such that, for every Hall set  $\mathcal{B} \subset \text{Br}(X)$ ,  $\mathbb{E}(\mathcal{A}) \not\subset \pm \mathbb{E}(\mathcal{B})$ .

*Proof.* The proof consists in constructing an example of a finite subset  $\mathcal{A}' \subset \text{Br}(X)$  such that  $\dim \text{span } \mathbb{E}(\mathcal{A}') = |\mathcal{A}'|$  (so  $\mathcal{A}'$  can be completed into an  $\mathcal{A}$  such that  $\mathbb{E}(\mathcal{A})$  is a basis of  $\mathcal{L}(X)$ ) and assume by contradiction that there exists a Hall set  $\mathcal{B} \subset \text{Br}(X)$  such that, for every  $a \in \mathcal{A}'$ , there exists  $b \in \mathcal{B}$  such that  $\mathbb{E}(a) = \pm \mathbb{E}(b)$ . We start by an elementary remark requiring three letters, before proving the result when  $|X| \geq 4$ , and eventually extending the result to  $|X| \in \{2, 3\}$ .

*Step 1:* We prove that, if  $[X_i, [X_j, X_k]] \in \pm \mathbb{E}(\mathcal{B})$ , where  $X_i, X_j, X_k$  are distinct elements of  $X$  and  $\mathcal{B}$  is a Hall set, then  $X_i \neq \min\{X_i, X_j, X_k\}$ , where the minimum is relative to the order in  $\mathcal{B}$ . Indeed, this implies that one of the following four cases occur.

- If  $(X_i, (X_j, X_k)) \in \mathcal{B}$ , then  $\lambda((X_j, X_k)) \leq X_i$ , so  $X_j \leq X_i$ , so  $X_j < X_i$  (since  $X_j \neq X_i$ ).
- If  $(X_i, (X_k, X_j)) \in \mathcal{B}$ , then  $\lambda((X_k, X_j)) \leq X_i$ , so  $X_k \leq X_i$ , so  $X_k < X_i$  (since  $X_k \neq X_i$ ).
- If  $((X_j, X_k), X_i) \in \mathcal{B}$ , then  $X_j = \lambda((X_j, X_k)) < (X_j, X_k) < X_i$ , so  $X_j < X_i$ .
- If  $((X_k, X_j), X_i) \in \mathcal{B}$ , then  $X_k = \lambda((X_k, X_j)) < (X_k, X_j) < X_i$ , so  $X_k < X_i$ .



Step 2: We prove that the result holds when  $|X| \geq 4$ . Let  $X_1, X_2, X_3, X_4 \in X$ . Define

$$\mathcal{A}' := \{(X_1, (X_2, X_3)), (X_2, (X_1, X_3)), (X_3, (X_4, X_1)), (X_4, (X_3, X_2))\}. \quad (\text{E.1})$$

Assume that there exists a Hall set  $\mathcal{B} \subset \text{Br}(X)$  such that  $\text{E}(\mathcal{A}') \subset \pm \text{E}(\mathcal{B})$ . Using the previous step and the first two brackets, we obtain that  $X_1 \neq \min\{X_1, X_2, X_3\}$  and  $X_2 \neq \min\{X_1, X_2, X_3\}$ , so  $X_3 < X_1$ . By symmetry, using the last two brackets,  $X_1 < X_3$ , which is a contradiction.

Step 3: Proof when  $|X| \in \{2, 3\}$ . Let  $X_0, X_1 \in X$ . For  $i \in \mathbb{N}$  let  $N_i := \text{ad}_{X_0}^i(X_1)$  and  $M_i := \text{ad}_{X_1}^i(X_0)$ . Define

$$\mathcal{A}'_N := \{(N_1, (N_2, N_3)), (N_2, (N_1, N_3)), (N_3, (N_4, N_1)), (N_4, (N_3, N_2))\}, \quad (\text{E.2})$$

$$\mathcal{A}'_M := \{(M_1, (M_2, M_3)), (M_2, (M_1, M_3)), (M_3, (M_4, M_1)), (M_4, (M_3, M_2))\} \quad (\text{E.3})$$

and  $\mathcal{A}' := \mathcal{A}'_N \cup \mathcal{A}'_M$ . Assume that there exists a Hall set  $\mathcal{B} \subset \text{Br}(X)$  such that  $\text{E}(\mathcal{A}') \subset \pm \text{E}(\mathcal{B})$ . Assume that, in  $\mathcal{B}$ ,  $X_0 < X_1$  (the other case being symmetric using  $M$  instead of  $N$ ). Since  $X_0 < X_1$ , for each  $i \in \mathbb{N}$ ,  $N_i \in \mathcal{B}$  and the set  $N := \{N_i; i \in \mathbb{N}\}$  is a free alphabetic subset of  $\mathcal{B}$ . By Proposition 3.5,  $\mathcal{B}_N := \tau^{-1}(\mathcal{B} \cap \text{Br}_N)$  is a Hall set of  $\text{Br}(N)$ , and the canonical morphism of Lie algebras  $\mathcal{L}(N) \rightarrow \mathcal{L}(X)$  induces an isomorphism onto the Lie subalgebra generated by  $N$ . Thus,  $\mathcal{B}_N \subset \text{Br}(N)$  is a Hall set such that  $\text{E}(\mathcal{A}'_N) \subset \text{E}(\mathcal{B}_N)$  with  $|N| \geq 4$ . This contradicts Step 2.  $\square$

## References

- [1] Luis Alvarez-Gaumé and Philippe Zaugg. Structure constants in the  $N = 1$  superoperator algebra. *Ann. Phys.*, 215(1):171–230, 1992.
- [2] Karine Beauchard, Jérémy Le Borgne, and Frédéric Marbach. On expansions for nonlinear systems, error estimates and convergence issues. *arXiv e-prints*, page arXiv:2012.15653, December 2020.
- [3] Dieter Blessenohl and Hartmut Laue. A basic construction for free Lie algebras. *Expo. Math.*, 11(2):145–152, 1993.
- [4] Nicolas Bourbaki. *Elements of mathematics. Algebra I. Chapters 1-3. Transl. from the French*. Springer-Verlag, 2nd printing edition, 1989.
- [5] Stephen Buckley, Djalil Chafai, Lajos Gyurko, Arend Janssen, and Terry Lyons. Libalgebra C++ Package, Computational Rough Paths. <https://sourceforge.net/projects/coropa/>, 2017.
- [6] Bill Casselman. Structure constants of Kac-Moody Lie algebras. In *Symmetry: representation theory and its applications. In honor of Nolan R. Wallach*, pages 55–83. New York, NY: Birkhäuser/Springer, 2014.
- [7] Bill Casselman. On Chevalley’s formula for structure constants. *J. Lie Theory*, 25(2):431–441, 2015.
- [8] Bill Casselman. Free lie algebras. <https://secure.math.ubc.ca/~cass/research/pdf/Free.pdf>, 2020. Last accessed 2021-07-17.
- [9] Bill Casselman. A simple way to compute structure constants of semi-simple lie algebras. *arXiv preprint arXiv:2011.08274*, 2020.
- [10] Kuo-Tsai Chen, Ralph Fox, and Roger Lyndon. Free differential calculus. IV: The quotient groups of the lower central series. *Ann. Math. (2)*, 68:81–95, 1958.

- [11] Claude Chevalley. Sur certains groupes simples. *Tohoku Math. J. (2)*, 7:14–66, 1955.
- [12] Evgenii Chibrikov. A right normed basis for free Lie algebras and Lyndon-Shirshov words. *J. Algebra*, 302(2):593–612, 2006.
- [13] Miriam Cohen and Sara Westreich. Structure constants related to symmetric Hopf algebras. *J. Algebra*, 324(11):3219–3240, 2010.
- [14] Michael Cuntz. Fusion algebras with negative structure constants. *J. Algebra*, 319(11):4536–4558, 2008.
- [15] Ignacy Duleba. Checking controllability of nonholonomic systems via optimal generation of Ph. Hall basis. *IFAC Proceedings Volumes*, 30(20):465–470, 1997.
- [16] Ignacy Duleba and Iwona Karcz-Duleba. Algorithm to Express Lie Monomials in Ph. Hall Basis and Its Practical Applications. In *International Conference on Computer Aided Systems Theory*, pages 465–473. Springer, 2019.
- [17] Carlos González. Dense orderings, partitions and weak forms of choice. *Fundam. Math.*, 147(1):11–25, 1995.
- [18] Henry Gould. Combinatorial identities. A standardized set of tables listing 500 binomial coefficient summations. Rev. ed. Morgantown, 1972.
- [19] Ronald Graham, Donald Knuth, and Oren Patashnik. *Concrete mathematics: a foundation for computer science*. Amsterdam: Addison-Wesley Publishing Group, 2nd edition, 1994.
- [20] Marshall Hall. A basis for free Lie rings and higher commutators in free groups. *Proc. Am. Math. Soc.*, 1:575–581, 1950.
- [21] Philip Hall. A contribution to the theory of groups of prime-power order. *Proc. Lond. Math. Soc. (2)*, 36:29–95, 1933.
- [22] Nathan Jacobson. *Lie algebras*. Dover Publications, Inc., New York, 1979. Republication of the 1962 original.
- [23] Matthias Kawski. Calculating the logarithm of the chen flies series. In *Proc. MTNS, Perpignan*, 2000.
- [24] Jean-Paul Laumond. Singularities and topological aspects in nonholonomic motion planning. In *Nonholonomic motion planning*, pages 149–199. Dordrecht: Kluwer Academic Publishers Group, 1992.
- [25] Michel Lazard. Groupes, anneaux de Lie et problème de Burnside. C.I.M.E., Gruppi, Anelli di Lie e Teoria della Coomologia, 1960.
- [26] M. Lothaire. *Combinatorics on words. Foreword by Roger Lyndon. 2nd ed*, volume 17. Cambridge: Cambridge University Press, 2nd ed. edition, 1997.
- [27] Roger Lyndon. On Burnside’s problem. *Trans. Am. Math. Soc.*, 77:202–215, 1954.
- [28] Heinrich Meier-Wunderli. Note on a basis of P. Hall for the higher commutators in free groups. *Commentarii Mathematici Helvetici*, 26(1):1–5, 1952.
- [29] Guy Melançon. Combinatorics of Hall trees and Hall words. *J. Comb. Theory, Ser. A*, 59(2):285–308, 1992.
- [30] Guy Melançon and Christophe Reutenauer. Lyndon words, free algebras and shuffles. *Can. J. Math.*, 41(4):577–591, 1989.

- [31] Duncan Rand, Pavel Winternitz, and Hans Zassenhaus. On the identification of a Lie algebra given by its structure constants. I: Direct decompositions, Levi decompositions, and nilradicals. *Linear Algebra Appl.*, 109:197–246, 1988.
- [32] Christophe Reutenauer. *Free Lie algebras*. Oxford: Clarendon Press, 1993.
- [33] Christophe Reutenauer. Free Lie algebras. In *Handbook of algebra. Volume 3*, pages 887–903. Amsterdam: Elsevier, 2003.
- [34] Herbert Robbins. A remark on Stirling’s formula. *Am. Math. Mon.*, 62:26–29, 1955.
- [35] Marcel-Paul Schützenberger. Sur une propriété combinatoire des demigroupes libres. *C. R. Acad. Sci., Paris*, 245:16–18, 1957.
- [36] Christoph Schwer. Galleries, Hall-Littlewood polynomials, and structure constants of the spherical Hecke algebra. *Int. Math. Res. Not.*, 2006(21):31, 2006. Id/No 75395.
- [37] Jean-Pierre Serre. *Free Lie Algebras*, pages 18–30. Springer Berlin Heidelberg, Berlin, Heidelberg, 1992.
- [38] Anatolii Shirshov. On free Lie rings. *Mat. Sb., Nov. Ser.*, 45:113–122, 1958.
- [39] Anatolii Shirshov. On the bases of a free Lie algebra. *Algebra Logika*, 1(1):14–19, 1962.
- [40] Stephen George Simpson. *The Hall Collection Process*. PhD thesis, Lehigh University, 1966.
- [41] Ralph Stöhr. Bases, filtrations and module decompositions of free Lie algebras. *J. Pure Appl. Algebra*, 212(5):1187–1206, 2008.
- [42] Hector Sussmann. A product expansion for the Chen series. Theory and applications of nonlinear control systems, Sel. Pap. 7th Int. Symp. Math. Theory Networks Syst., Stockholm, 1986.
- [43] Alberto Tagiuri. Di alcune successioni ricorrenti a termini interi e positivi. *Periodico di Mat.*, 16(1):1–12, 1901.
- [44] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 9.3)*, 2021. <https://www.sagemath.org>.
- [45] Jacques Tits. Sur les constantes de structure et le théorème d’existence des algèbres de Lie semi-simples. *Publ. Math., Inst. Hautes Étud. Sci.*, 31:525–562, 1966.
- [46] Miguel Torres-Torriti and Hannah Michalska. A software package for Lie algebraic computations. *SIAM Rev.*, 47(4):722–745, 2005.
- [47] Nikolai Vavilov. Can one see the signs of structure constants? *St. Petersburg Math. J.*, 19(4):519–543, 2008.
- [48] Gérard Viennot. *Algèbres de Lie libres et monoïdes libres*, volume 691 of *Lecture Notes in Mathematics*. Springer, Berlin, 1978. Bases des algèbres de Lie libres et factorisations des monoïdes libres.
- [49] Gérard Viennot. Quelques bases et familles basiques des algèbres de Lie libres commodes pour les calculs sur ordinateurs. *Bull. Soc. Math. Fr., Suppl., Mém.*, 49-50:201–209, 1977.
- [50] Ernst Witt. Die Unterringe der freien Lieschen Ringe. *Math. Z.*, 64:195–216, 1956.