



**HAL**  
open science

# Lower Bounds for Arithmetic Circuits via the Hankel Matrix

Nathanaël Fijalkow, Guillaume Lagarde, Pierre Ohlmann, Olivier Serre

► **To cite this version:**

Nathanaël Fijalkow, Guillaume Lagarde, Pierre Ohlmann, Olivier Serre. Lower Bounds for Arithmetic Circuits via the Hankel Matrix. *Computational Complexity*, 2021, 30 (2), 10.1007/s00037-021-00214-1. hal-03383188

**HAL Id: hal-03383188**

**<https://hal.science/hal-03383188v1>**

Submitted on 18 Oct 2021

**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.

# LOWER BOUNDS FOR ARITHMETIC CIRCUITS VIA THE HANKEL MATRIX

NATHANAËL FIJALKOW, GUILLAUME LAGARDE,  
PIERRE OHLMANN, AND OLIVIER SERRE

March 12, 2021

**Abstract.** We study the complexity of representing polynomials by arithmetic circuits in both the commutative and the non-commutative settings. Our approach goes through a precise understanding of the more restricted setting where multiplication is not associative, meaning that we distinguish  $(xy)z$  from  $x(yz)$ .

Our first and main conceptual result is a characterization result: we show that the size of the smallest circuit computing a given non-associative polynomial is exactly the rank of a matrix constructed from the polynomial and called the Hankel matrix. This result applies to the class of all circuits in both commutative and non-commutative settings, and can be seen as an extension of the seminal result of Nisan giving a similar characterization for non-commutative algebraic branching programs.

The study of the Hankel matrix provides a unifying approach for proving lower bounds for polynomials in the (classical) associative setting. Our key technical contribution is to provide generic lower bound theorems based on analyzing and decomposing the Hankel matrix. We obtain significant improvements on lower bounds for circuits with many parse trees, in both (associative) commutative and non-commutative settings, as well as alternative proofs of recent results proving superpolynomial and exponential lower bounds for different classes of circuits as corollaries of our characterization and decomposition results.

**Keywords.** Arithmetic Circuit Complexity, Lower Bounds, Parse Trees, Hankel Matrix

**Subject classification.** 68Q17

# 1. Introduction

The model of arithmetic circuits is the algebraic analogue of Boolean circuits: the latter computes Boolean functions and the former computes polynomials, replacing OR gates by addition and AND gates by multiplication. Computational complexity theory is concerned with understanding the expressive power of such models. A rich theory investigates the algebraic complexity classes **VP** and **VNP** introduced by Valiant (Valiant 1979). A widely open problem in this area of research is to explicitly construct hard polynomials, meaning for which we can prove superpolynomial lower bounds. To this day the best general lower bounds for arithmetic circuits were given by Baur and Strassen (Baur & Strassen 1983) for the polynomial  $\sum_{i=1}^n x_i^d$ , which requires  $\Omega(n \log d)$  operations.

The seminal paper of Nisan (Nisan 1991) initiated the study of non-commutative computation: in this setting variables do not commute, and therefore  $xy$  and  $yx$  are considered as being two distinct monomials. Non-commutative computations arise in different scenarios, the most common mathematical examples being when working with algebras of matrices, group algebras of non-commutative groups or the quaternion algebra. A second motivation for studying the non-commutative setting is that it makes it easier to prove lower bounds which can then provide powerful ideas for the commutative case. Indeed, commutativity allows a circuit to rely on cancellations and to share calculations across different gates, making them more complicated to analyze.

**1.1. Nisan’s Characterization for ABP.** The main result of Nisan (Nisan 1991) is to give a characterization of the smallest ABP computing a given polynomial. As a corollary of this characterization Nisan obtains exponential lower bounds for the non-commutative permanent against the subclass of circuits given by ABPs.

We sketch the main ideas behind Nisan’s characterization, since our first contribution is to extend these ideas to the class of all non-associative circuits. An ABP is a layered graph with two distinguished vertices, a source and a target. The edges are labelled by affine functions in a given set of variables. An ABP computes a

46 polynomial obtained by summing over all paths from the source to  
 47 the target, with the value of a path being the multiplication of the  
 48 affine functions along the traversed edges. Following Nisan, fix a  
 49 polynomial  $f$ , and define a matrix  $N_f$  whose rows and columns are  
 50 indexed by monomials: for  $u, v$  two monomials, let  $N_f(u, v)$  denote  
 51 the coefficient of the monomial  $u \cdot v$  in  $f$ .

52 The beautiful and surprisingly simple characterization of Nisan  
 53 states that for a homogeneous (i.e., all monomials have the same  
 54 degree) non-commutative polynomial  $f$ , the size of the smallest  
 55 ABP computing  $f$  is exactly the rank of  $N_f$ . The key idea is to  
 56 decompose the computation arising in the ABP, say  $\mathcal{C}$ : to any  
 57 vertex  $r$  in  $\mathcal{C}$  we associate two polynomials  $L_r$  and  $R_r$  that are  
 58 respectively the one computed by the ABP induced by the original  
 59 source of  $\mathcal{C}$  and target  $r$  and the one computed by the ABP induced  
 60 by source  $r$  and the original target of  $\mathcal{C}$ . For a polynomial  $f$  and  
 61 a monomial  $m$  we use  $f(m)$  to denote the coefficient of  $m$  in  $f$ .  
 62 For  $u, v$  two monomials, we observe that the coefficient of  $u \cdot v$  in  
 63  $f$  is equal to  $\sum_r L_r(u)R_r(v)$ , where  $r$  ranges over all vertices of  
 64  $\mathcal{C}$ ,  $L_r(u)$  is the coefficient of  $u$  in  $L_r$ , and  $R_r(v)$  is the coefficient  
 65 of  $v$  in  $R_r$ . We see this as a matrix equality:  $N_f = \sum_r L_r \cdot R_r$ ,  
 66 where  $L_r$  is seen as a column vector, and  $R_r$  as a row vector. By  
 67 subadditivity of the rank and since the product of a column vector  
 68 by a row vector is a matrix of rank at most 1, this implies that the  
 69 rank of  $N_f$  is bounded by the size of the ABP, yielding the lower  
 70 bound in Nisan's result.

71 The crucial idea of splitting the computation of a monomial into  
 72 two parts had been independently developed by Fliess when study-  
 73 ing so-called *Hankel Matrices* in (Fliess 1974) to derive a very sim-  
 74 ilar result in the field of *weighted automata*, which are finite state  
 75 machines recognising *words series*, i.e., functions from finite words  
 76 into a field. Fliess' theorem (Fliess 1974, Th. 2.1.1) states that the  
 77 size of the smallest weighted automaton recognising a word series  
 78  $f$  is exactly the rank of the Hankel matrix of  $f$ . The key insight to  
 79 relate the two results is to see a non-commutative monomial as a  
 80 finite word over the alphabet whose letters are the variables. Using  
 81 this correspondence one can obtain Nisan's theorem from Fliess'  
 82 theorem, observing that the Hankel matrix coincides with the ma-

83 trix  $N_f$  defined by Nisan and that acyclic weighted automata corre-  
 84 spond to ABPs. (We refer to an early technical report of this work  
 85 for more details on this correspondence (Fijalkow *et al.* 2018).)

86 **1.2. Non-Associative Computations.** Hrubeš, Wigderson and  
 87 Yehudayoff (Hrubeš *et al.* 2011) drop the associativity rule and  
 88 show how to define the complexity classes **VP** and **VNP** in the ab-  
 89 sence of either commutativity or associativity (or both) and prove  
 90 that these definitions are sound in particular by obtaining the com-  
 91 pleteness of the permanent.

92 In the same way that a non-commutative monomial can be seen  
 93 as a word, a non-commutative and non-associative monomial such  
 94 as  $(xy)(x(zy))$  can be seen as a tree, and more precisely as an or-  
 95 dered binary rooted tree whose leaves are labelled by variables. The  
 96 starting point of our work was to exploit this connection. The work  
 97 of Bozapalidis and Louscou-Bozapalidou (Bozapalidis & Louscou-  
 98 Bozapalidou 1983) extends Fliess’ result to trees; although we do  
 99 not technically rely on their results, they serve as a guide, in par-  
 100 ticular for understanding how to decompose trees.

101 Let us return to the key idea in Nisan’s proof, which is to  
 102 decompose the computation of an ABP into two parts. The way  
 103 a monomial, e.g.,  $x_1x_2x_3 \cdots x_d$ , is evaluated in an ABP is very  
 104 constrained, namely from left to right, or if we make the implicit  
 105 non-associative structure explicit as  $w = (\cdots (((x_1x_2)x_3)x_4) \cdots)x_d$ .  
 106 The decompositions of  $w$  into two monomials  $u, v$  are of the form  
 107  $u = (\cdots ((x_1x_2)x_3) \cdots)x_{i-1}$  and  $v = (\cdots ((\square x_i)x_{i+1}) \cdots)x_d$ . Here  
 108  $\square$  is a new fresh variable (the *hole*) to be substituted by  $u$ . Moving  
 109 to non-associative polynomials, a monomial is a tree whose leaves  
 110 are labelled by variables. A *context* is a monomial over the set of  
 111 variables extended with a new fresh one denoted  $\square$  and occurring  
 112 exactly once. For instance (see Figure 1.1) the composition of the  
 113 monomial  $t = z((xx)y)$  with the context  $c = (xy)((z\square)y)$  is the  
 114 monomial  $c[t] = (xy)((z(z((xx)y)))y)$ .

115 Let  $f$  be a non-associative (possibly commutative) polynomial  
 116  $f$ , the *Hankel matrix*  $H_f$  of  $f$  is defined as follows: the rows of  
 117  $H_f$  are indexed by contexts and the columns by monomials, and  
 118 the value of  $H_f(c, t)$  at row  $c$  and column  $t$  is the coefficient of the  
 119 monomial  $c[t]$  in  $f$ .

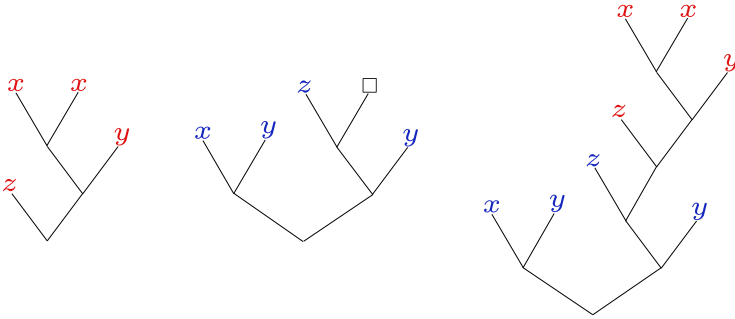


Figure 1.1: On the left hand side the monomial  $t$ , in the middle the context  $c$ , and on the right hand side the monomial  $c[t]$ .

120        Extending Nisan’s proof to computations in a *general circuit*,  
 121 which are done along trees, we obtain a characterization in the  
 122 non-associative setting (a more precise statement is given by The-  
 123 orem 2.4)).

124 **THEOREM.** *Let  $f$  be a non-associative homogeneous polynomial*  
 125 *and let  $H_f$  be its Hankel matrix. Then, the size of the smallest*  
 126 *circuit computing  $f$  is exactly  $\text{rank}(H_f)$ .*

127        Note that this is a characterization result: the Hankel matrix  
 128 exactly captures the size of the smallest circuit computing  $f$  (up-  
 129 per and lower bounds), exactly as in Nisan’s result. Hence, under-  
 130 standing the rank of the Hankel matrix is equivalent to studying  
 131 circuits for  $f$ . We recover and extend Nisan’s characterization as  
 132 a special case of our result.

133 **1.3. Parse Trees.** At an intuitive level, parse trees can be used  
 134 to explain in what way a circuit uses the associativity rule. Con-  
 135 sider the case of a circuit computing the (associative) monomial  
 136  $2xyz$ . Since this monomial corresponds to two non-associative  
 137 monomials:  $(xy)z$  and  $x(yz)$ , the circuit may sum different com-  
 138 putations, for instance  $3(xy)z - x(yz)$ , which up to associativity is  
 139  $2xyz$ . We say that such a circuit contains two parse trees, corre-  
 140 sponding to the two different ways of parenthesizing  $xyz$ .

141        The *shape* of a non-associative monomial is the tree obtained  
 142 by forgetting the variables, e.g., the shape of  $(z((xy)((xx)y)))$  is

143  $(\_ ((\_ \_)((\_ \_)\_)))$ . The parse trees of a circuit  $\mathcal{C}$  are the shapes  
 144 induced by computations in  $\mathcal{C}$ .

145 Many interesting classes of circuits can be defined by restricting  
 146 the set of allowed parse trees, both in the commutative and the  
 147 non-commutative setting.

148 ○ The simplest such class is that of Algebraic Branching Pro-  
 149 grams (ABP) (Nisan 1991; Dvir *et al.* 2012; Ramya & Rao  
 150 2018), whose only parse trees are left-combs, that is, the vari-  
 151 ables are multiplied sequentially.

152 ○ Lagarde, Malod and Perifel (Lagarde *et al.* 2016) introduced  
 153 the class of Unique Parse Tree circuits (UPT), which are  
 154 circuits computing non-commutative homogeneous (but as-  
 155 sociative) polynomials such that all monomials are evaluated  
 156 in the same non-associative way.

157 ○ The class of skew circuits (Toda 1992; Allender *et al.* 1998;  
 158 Malod & Portier 2008; Limaye *et al.* 2016) and its exten-  
 159 sion to small non-skew depth circuits (Limaye *et al.* 2016),  
 160 together with the class of unambiguous circuits (Arvind &  
 161 Raja 2016) are all defined via parse tree restrictions.

162 ○ We propose in our technical developments some related re-  
 163 strictions called slightly balanced and slightly unbalanced cir-  
 164 cuits.

165 ○ Last but not least, the class of  $k$ -PT circuits (Arvind & Raja  
 166 2016; Saptharishi & Tengse 2017; Lagarde *et al.* 2018) is sim-  
 167 ply the class of circuits having at most  $k$  parse trees.

168 **1.4. Contributions and Outline.** In this paper we prove lower  
 169 bounds for classes of circuits with parse tree restrictions, both in  
 170 the commutative and non-commutative setting.

171 Our first and conceptually main contribution is the character-  
 172 ization result stated in Theorem 2.4 which gives an algebraic ap-  
 173 proach to understanding circuits in the non-associative setting. All  
 174 the subsequent results in this paper are based on this approach.

175 Section 3.1 and Section 3.2 are devoted to the definition of parse  
 176 trees and a classical tool for proving lower bounds, partial deriva-  
 177 tive matrices. We can already show at this point (in Section 3.3)  
 178 how Theorem 2.4 can be specialized to give a characterization re-  
 179 sult for UPT circuits, extending Nisan’s result. We note that a  
 180 characterization result for UPT circuits was already known (La-  
 181 garde *et al.* 2016), we slightly improve on it. As a corollary we  
 182 obtain exponential lower bounds on the size of the smallest UPT  
 183 circuit computing the permanent.

184 Our most technical developments are discussed in Section 4.  
 185 We prove generic lower bound results by further analyzing and  
 186 decomposing the Hankel matrix, with the following proof scheme.  
 187 We consider a polynomial  $f$  in the associative setting. Let  $\mathcal{C}$  be a  
 188 circuit computing  $f$ . Forgetting about associativity we can see  $\mathcal{C}$  as  
 189 computing a non-associative polynomial  $\tilde{f}$ , which projects onto  $f$ ,  
 190 meaning is equal to  $f$  assuming associativity. This induces a set of  
 191 linear constraints: for instance if the monomial  $xyz$  has coefficient  
 192 3 in  $f$ , then we know that  $\tilde{f}((xy)z) + \tilde{f}(x(yz)) = 3$ . We make use  
 193 of the linear constraints to derive lower bounds on the rank of the  
 194 Hankel matrix  $H_{\tilde{f}}$ , yielding a lower bound on the size of  $\mathcal{C}$ .

195 The final section is devoted to applications of our results, where  
 196 we obtain superpolynomial and exponential lower bounds for var-  
 197 ious classes. In the results mentioned below,  $n$  is the number of  
 198 variables,  $d$  is the degree of the polynomial, and  $k$  the number of  
 199 parse trees. We note that the lower bounds hold for any (prime)  
 200  $n$ , any  $d$ , and any field.

201 We obtain alternative proofs of some known lower bounds: un-  
 202 ambiguous circuits (Arvind & Raja 2016), skew circuits (Limaye  
 203 *et al.* 2016) and small non-skew depth circuits (obtaining a much  
 204 shorter proof than (Limaye *et al.* 2016)).

205 Our novel results are:

- 206 ○ *Slightly unbalanced circuits.* We extend the exponential lower  
 207 bound from (Limaye *et al.* 2016) on  $\frac{1}{5}$ -unbalanced circuits to  
 208  $(\frac{1}{2} - \varepsilon)$ -unbalanced circuits.
- 209 ○ *Slightly balanced circuits.* We derive a new exponential lower  
 210 bound for  $\varepsilon$ -balanced circuits.



- 211 ○ *Circuits with  $k$  parse trees in the non-commutative setting.*  
 212 We substantially extend the superpolynomial lower bound  
 213 of (Lagarde *et al.* 2018) from  $k = 2^{d^{1/3-\varepsilon}}$  to  $k = 2^{d^{1-\varepsilon}}$ , the  
 214 total number of possible non-commutative parse trees being  
 215  $2^{O(d)}$ .
- 216 ○ *Circuits with  $k$  parse trees in the commutative setting.* We  
 217 substantially extend the superpolynomial lower bound from (Arvind  
 218 & Raja 2016) from  $k = d^{1/2-\varepsilon}$  to  $k = 2^{d^{1/3-\varepsilon}}$ , and even to  
 219  $k = 2^{d^{1-\varepsilon}}$ , when  $d$  is polylogarithmic in  $n$ .

220 **1.5. Related Work.** We argued that proving lower bounds in  
 221 the non-commutative setting is easier, but this has not yet ma-  
 222 terialized since the best lower bound for general circuits in this  
 223 setting is the same as in the commutative setting (by Baur and  
 224 Strassen, already mentioned above). Indeed, recent impressive  
 225 results suggest that this may be hard: Carmosino, Impagliazzo,  
 226 Lovett, and Mihajlin (Carmosino *et al.* 2018) (essentially) proved  
 227 that a lower bound in the non-commutative setting which would  
 228 be slightly stronger than superlinear can be amplified to get strong  
 229 lower bounds (even exponential, in some cases) again in the non-  
 230 commutative setting.

231 Most approaches for proving lower bounds rely on algebraic  
 232 techniques and the rank of some matrix. A different and beautiful  
 233 approach was investigated by Hrubeš, Wigderson and Yehuday-  
 234 off (Hrubeš *et al.* 2011) in the non-commutative setting through  
 235 the study of the so-called *sum-of-squares problem*. Roughly speak-  
 236 ing, the goal is to decompose  $(x_1^2 + \dots + x_k^2) \cdot (y_1^2 + \dots + y_k^2)$  into a sum  
 237 of  $n$  squared bilinear forms in the variables  $x_i$  and  $y_j$ . They show  
 238 that almost any superlinear bound on  $n$  implies non-trivial lower  
 239 bounds on the size of any non-commutative circuit computing the  
 240 permanent.

241 The quest of finding lower bounds is deeply connected to an-  
 242 other problem called polynomial identity testing (PIT) for which  
 243 the goal is to decide whether a given circuit computes the formal  
 244 zero polynomial. The connection was shown in (Kabanets & Im-  
 245 pagliazzo 2003), in which it is proved that providing an efficient  
 246 deterministic algorithm to solve the problem implies strong lower

247 bounds either in the arithmetic or boolean setting. PIT was widely  
 248 investigated in the commutative and non-commutative settings for  
 249 classes of circuits based on parse trees restrictions, see e.g., (Raz &  
 250 Shpilka 2005; Forbes *et al.* 2014; Agrawal *et al.* 2015; Gurjar *et al.*  
 251 2017; Saptharishi & Tengse 2017; Arvind *et al.* 2017).

## 252 2. Characterizing Non-Associative Circuits

253 **2.1. Basic Definitions.** For an integer  $d \in \mathbb{N}$ , we let  $[d]$  denote  
 254 the integer interval  $\{1, \dots, d\}$ .

255 **Polynomials.** Let  $K$  be a field and let  $X$  be a set of *variables*.  
 256 Following (Hrubeš *et al.* 2011) we consider that unless otherwise  
 257 stated multiplication is neither commutative nor associative. We  
 258 assume however that addition is commutative and associative, and  
 259 that multiplication distributes over addition. A *monomial* is a  
 260 product of variables in  $X$  and a polynomial  $f$  is a formal finite  
 261 sum  $\sum_i c_i m_i$  where  $m_i$  is a monomial and  $c_i \in K$  is a non-zero  
 262 element called the coefficient of  $m_i$  in  $f$ . We let  $f(m_i)$  denote the  
 263 coefficient of  $m_i$  in  $f$ , so that  $f = \sum_i f(m_i) m_i$ .

264 The *degree* of a monomial is defined in the usual way, i.e.,  
 265  $\deg(x) = 1$  when  $x \in X$  and  $\deg(m_1 m_2) = \deg(m_1) + \deg(m_2)$ ; the  
 266 degree of a polynomial  $f$  is the maximal degree of a monomial in  $f$ .  
 267 A polynomial is *homogeneous* if all its monomials have the same  
 268 degree. Depending on whether we include the relations  $u \cdot v = v \cdot u$   
 269 (commutativity) and  $u \cdot (v \cdot w) = (u \cdot v) \cdot w$  (associativity) we obtain  
 270 four classes of polynomials.

271 Unless otherwise specified, for a polynomial  $f$  we use  $n$  for the  
 272 number of variables and  $d$  for the degree.

273 **Trees and Contexts.** The *trees* we consider have a single root  
 274 and binary branching (every internal node has exactly two chil-  
 275 dren). To account for the commutative and for the non-commutative  
 276 setting we use either *unordered trees* or *ordered trees*, the only  
 277 difference being that in the case of ordered trees we distinguish the  
 278 left child from the right child. We let *Tree* denote the set of trees  
 279 (it will be clear from the context whether they are ordered or not).  
 280 The size of a tree is defined as its number of leaves.

281 A **non-associative monomial**  $t$  is a tree with leaves labelled  
 282 by variables. If  $t$  is non-commutative then it is an ordered tree, and  
 283 if  $t$  is commutative then it is an unordered tree. We let  $\mathbf{Tree}(X)$   
 284 denote the set of trees whose leaves are labelled by variables in  
 285  $X$  and  $\mathbf{Tree}_i(X)$  denote the subset of such trees with  $i$  leaves,  
 286 which are monomials of degree  $i$ . Given a non-associative mono-  
 287 mial  $t$ , we let  $\text{label}(t)$  be the associative monomial corresponding  
 288 to the multiplication of the variables at the leaves of  $t$ . If  $t$  is non-  
 289 commutative, the multiplication is done from left to right, and  
 290  $\text{label}(t)$  is a non-commutative monomial, that is, a word.

291 In this paper, we see polynomials as finitely supported map-  
 292 pings from monomials to  $K$ . For instance, in the non-associative  
 293 setting where monomials are trees, a non-associative polynomial is  
 294 a map  $\mathbf{Tree}(X) \rightarrow K$ . To avoid possible confusion, let us insist  
 295 that the notation  $f(t)$  refers to the coefficient of the monomial  $t$  in  
 296 the polynomial  $f$ , not to be confused with the evaluation of  $f$  at a  
 297 given point.

298 A (ordered or unordered) **context** is a tree with a distinguished  
 299 leaf labelled by a special symbol called the **hole** and written  $\square$ .  
 300 We let  $\mathbf{Context}(X)$  denote the set of contexts whose leaves are  
 301 labelled by variables in  $X$ . Given a context  $c$  and a tree  $t$  we  
 302 construct a new tree  $c[t]$  by substituting the hole of  $c$  by  $t$ . This  
 303 operation is defined in both ordered and unordered settings. See  
 304 Figure 1.1 for an example. It can be read in both the ordered or  
 305 unordered settings.

306 **Hankel Matrices.** Let  $f$  be a non-associative polynomial. The  
 307 **Hankel matrix**  $H_f$  of  $f$  is the matrix whose rows are indexed by  
 308 contexts and columns by monomials and such that the value of  $H_f$   
 309 at row  $c$  and column  $t$  is the coefficient of the monomial  $c[t]$  in  $f$ .  
 310 Note that  $H_f$  is an infinite matrix with finite support, so its rank  
 311 is well defined. As we will be interested in computing the rank of  
 312  $H_f$ , we freely depict its rows and columns ordered arbitrarily and  
 313 conveniently.

314 **Arithmetic Circuits.** An (arithmetic) **circuit** is a directed acyclic  
 315 graph such that the vertices are of three types:

316   ◦ input gates: they have in-degree 0 and are labelled by vari-

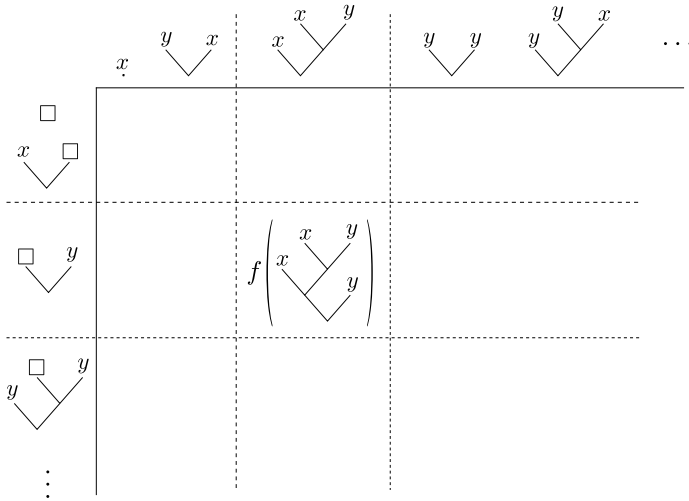


Figure 2.1: A depiction of the Hankel matrix of a non-associative polynomial  $f$ . Only one coefficient is displayed for clarity.

- 317       ables in  $X$ ,
- 318       ◦ addition gates: they have arbitrary in-degree, an output
- 319       weight in  $K$ , and a weight  $w(a) \in K$  on each incoming arc  $a$ ,
- 320       ◦ multiplication gates: they have in-degree 2, and we distin-
- 321       guish between the left child and the right child.

322 Each gate  $v$  in the circuit computes a polynomial  $f_v$  which we

323 define by induction.

- 324       ◦ An input gate labelled by a variable  $x \in X$  computes the
- 325       polynomial  $x$ .
- 326       ◦ An addition gate  $v$  with  $n$  arcs incoming from gates  $v_1, \dots, v_n$
- 327       and with weights  $\alpha_1, \dots, \alpha_n$ , computes the polynomial  $\alpha_1 f_{v_1} +$
- 328        $\dots + \alpha_n f_{v_n}$ .
- 329       ◦ A multiplication gate with left child  $u$  and right child  $v$  com-
- 330       putes the polynomial  $f_u f_v$ .

331 The circuit itself computes a polynomial given by the sum over

332 all addition gates of the polynomial computed by the gate times

333 its output weight. Note that it is slightly unusual that all addition  
 334 gates contribute to the circuit; one can easily reduce to the classical  
 335 case where there is a unique output addition gate by adding an  
 336 extra gate.

337 We shall make a syntactic assumption: each arc is either coming  
 338 from, or going to (but not both), an addition gate. Any circuit can  
 339 be put into this form by adding addition gates, at most one per  
 340 input gate and per multiplication gate (see Figure 2.2). We also  
 341 ask two input gates referring to the same variable to not feed the  
 342 same addition gate. We then define the size of a circuit to be its  
 343 number of addition gates, which compensates this small blow up.  
 344 Doing so we slightly differ from usual, however this will allow our  
 345 characterization result to be exact.

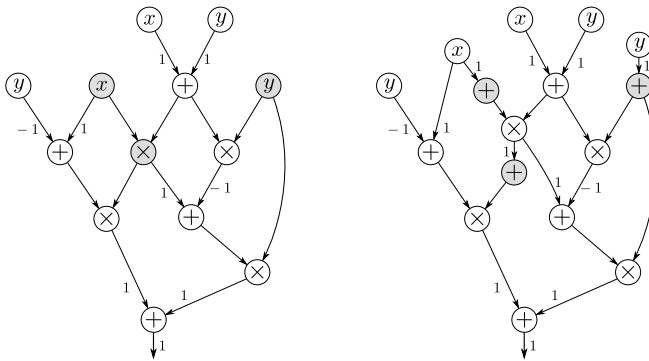
346 Note that the definitions we gave above do not depend on which  
 347 of the four settings we consider: commutative or non-commutative,  
 348 associative or non-associative.

349 **2.2. The Characterization.** This section aims at proving the  
 350 characterization stated in Theorem 2.4 below — the Hankel ma-  
 351 trix  $H_f$  exactly captures (upper and lower bounds) the size of  
 352 the smallest circuit computing  $f$  —, extending Nisan’s character-  
 353 ization of non-commutative ABPs to general circuits in the non-  
 354 associative setting. The result holds for both commutative and  
 355 non-commutative settings, the proof being the same up to cosmetic  
 356 changes.

357 The key step to go from ABPs to general circuits is the fol-  
 358 lowing: the polynomial computed by an ABP is the sum over the  
 359 *paths* of the underlying graph, whereas in a general circuit the sum  
 360 is over *trees*. We formalize this in the next definition by introducing  
 361 *runs* of a circuit. The definition is given in the non-commutative  
 362 setting but easily adapts to the commutative setting as explained  
 363 later in Remark 2.2.

364 **DEFINITION 2.1.** *Let  $\mathcal{C}$  be a circuit and  $V_{\oplus}$  denote its set of addi-*  
 365 *tion gates. Let  $t \in \text{Tree}(X)$  be a monomial. A **run of  $\mathcal{C}$  over  $t$***   
 366 *is a map  $\rho$  from nodes of  $t$  to  $V_{\oplus}$  such that*

- 367 (i) *A leaf of  $t$  with label  $x \in X$  is mapped to a gate with a*  
 368 *non-zero edge incoming from an input gate labelled by  $x$ .*



	A	NA
C	$x^3 + x^2y - y^2x - y^3$	
NC	$x^3 + 2x^2y - yx^2 - yxy - y^3$	

Figure 2.2: The circuit on the left does not satisfy our syntactic assumption because of the edges leaving the greyed gates. However, the one on the right, obtained by adding two addition gates does satisfy the assumption. It has size 6. Both circuit compute the same polynomials in each setting, which are given in the table below, where the abbreviations A, NA, C, NC respectively stand for associative, non-associative, commutative, non-commutative. We use labelled outgoing edges to depict output weights, and omit them when the output weight is 0.

369 (ii) If  $n$  is a node of  $t$  with left child  $n_1$  and right child  $n_2$ , then  
 370  $\rho(n)$  has a non-zero edge incoming from a multiplication gate  
 371 with left child  $\rho(n_1)$  and right child  $\rho(n_2)$ .

372 The **value**  $\text{val}(\rho)$  of  $\rho$  is a non-zero element in  $K$  defined as the  
 373 product of the weights of the edges mentioned in items (i) and (ii)  
 374 together with the output weight of  $\rho(r)$ ,  $r$  being the root of  $t$ .

375 We write by a slight abuse of notation  $\rho : t \rightarrow V_{\oplus}$  for runs of  $\mathcal{C}$   
 376 over  $t$ .

377 Figure 2.3 depicts a run in the circuit from Figure 2.2.

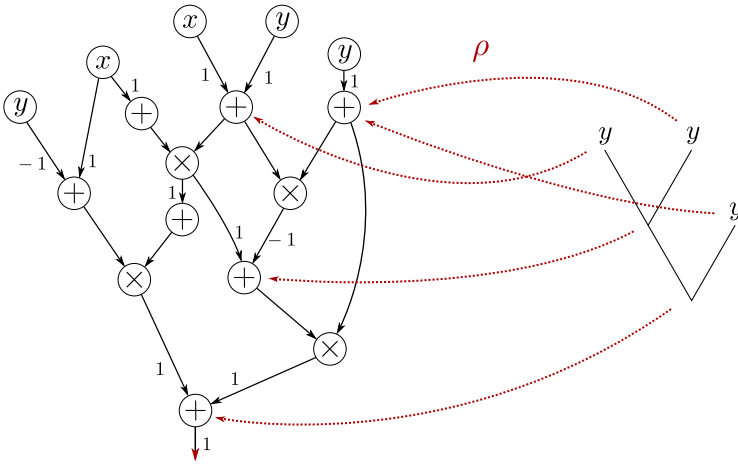


Figure 2.3: A run  $\rho$  in the circuit on the left, over the monomial on the right. It has value  $-1$ .

378 REMARK 2.2. In the commutative setting we simply replace item  
 379 (ii) by: “if  $n$  is a node of  $t$  with children  $n_1, n_2$ , then  $\rho(n)$  has a  
 380 non-zero edge incoming from a multiplication gate with children  
 381  $\rho(n_1), \rho(n_2)$ ”.

382 A run of  $\mathcal{C}$  over a monomial  $t$  additively contributes to the  
 383 coefficient of  $t$  in the polynomial computed by  $\mathcal{C}$ , leading to the  
 384 following straightforward lemma.

385 LEMMA 2.3. Let  $\mathcal{C}$  be a circuit computing the non-associative poly-  
 386 nomial  $f : \text{Tree}(X) \rightarrow K$ . Then the coefficient  $f(t)$  of a monomial  
 387  $t \in \text{Tree}(X)$  in  $f$  is equal to

$$\sum_{\rho: t \rightarrow V_{\oplus}} \text{val}(\rho).$$

388 We may now state and prove our cornerstone result, which holds  
 389 in both the commutative and non-commutative settings.

390 THEOREM 2.4. Let  $f : \text{Tree}(X) \rightarrow K$  be a non-associative poly-  
 391 nomial,  $H_f$  be its Hankel matrix, and  $\mathcal{C}$  be a circuit computing  
 392  $f$ . Then  $|\mathcal{C}| \geq \text{rank}(H_f)$ . Moreover, if  $f$  is homogeneous this  
 393 bound is tight, meaning there exists a circuit  $\mathcal{C}$  computing  $f$  of  
 394 size  $\text{rank}(H_f)$ .

395 Note that an interesting feature of this theorem is that the  
 396 upper bound is effective: given a homogenous polynomial one can  
 397 construct a circuit computing this polynomial of size  $\text{rank}(H_f)$ .

398 The proof of the lower bound follows the same lines as Nisan's  
 399 original proof for non-commutative ABPs (Nisan 1991).

400 PROOF. We start with the lower bound, that is,  $|\mathcal{C}| \geq \text{rank}(H_f)$ .

401 Let  $\mathcal{C}$  be a circuit computing the non-associative polynomial  
 402  $f : \text{Tree}(X) \rightarrow K$ . Let  $V_{\oplus}$  denote the set of addition gates of  $\mathcal{C}$ .  
 403 To bound the rank of the Hankel matrix  $H_f$  by  $|\mathcal{C}| = |V_{\oplus}|$  we show  
 404 that  $H_f$  can be written as the sum of  $|V_{\oplus}|$  matrices each of rank  
 405 at most 1.

406 For each  $v \in V_{\oplus}$  we define two circuits which decompose the  
 407 computations around  $v$ . Let  $\mathcal{C}_1^v$  be the circuit obtained from  $\mathcal{C}$  by  
 408 changing all output weights to 0 except that of  $v$  which is set to 1.  
 409 Note that  $\mathcal{C}_1^v$  can be seen as the restriction of  $\mathcal{C}$  to descendants of  $v$ .  
 410 Let  $\mathcal{C}_2^v$  be another copy of  $\mathcal{C}$  with just one extra input gate labelled  
 411 by a fresh variable  $\square \notin X$  with a single outgoing edge with weight  
 412 1 going to  $v$ . We let  $f^v : \text{Tree}(X) \rightarrow K$  denote the polynomial  
 413 computed by  $\mathcal{C}_1^v$  and  $g^v : \text{Context}(X) \rightarrow K$  denote the restriction  
 414 of the polynomial computed by  $\mathcal{C}_2^v$  to  $\text{Context}(X) \subseteq \text{Tree}(X \sqcup \{\square\})$ .

415 We now show the equality

$$H_f(c, t) = \sum_{v \in V_{\oplus}} f^v(t)g^v(c).$$

416 For that, fix a monomial  $t \in \text{Tree}(X)$  and a context  $c \in$   
 417  $\text{Context}(X)$  and denote by  $n_{\square}$  the leaf of  $c$  labelled by  $\square$ , which  
 418 is also the root of  $t$  and the node to which  $t$  is substituted with in  
 419  $c[t]$ . Relying on Lemma 2.3, we calculate the coefficient  $f(c[t])$  of



420  $c[t]$  in  $f$ .

$$\begin{aligned}
f(c[t]) &= \sum_{\rho: c[t] \rightarrow V_{\oplus}} \text{val}(\rho) = \sum_{v \in V_{\oplus}} \sum_{\substack{\rho: c[t] \rightarrow V_{\oplus} \\ \rho(n_{\square})=v}} \text{val}(\rho) \\
&= \sum_{v \in V_{\oplus}} \sum_{\substack{\rho_1^v: t \rightarrow V_{\oplus} \\ \rho_1^v(n_{\square})=v}} \sum_{\substack{\rho_2^v: c \rightarrow V_{\oplus} \\ \rho_2^v(n_{\square})=v}} \text{val}(\rho_1^v) \text{val}(\rho_2^v) \\
&= \sum_{v \in V_{\oplus}} \sum_{\substack{\rho_1^v: t \rightarrow V_{\oplus} \\ \rho_1^v(n_{\square})=v}} \text{val}(\rho_1^v) \sum_{\substack{\rho_2^v: c \rightarrow V_{\oplus} \\ \rho_2^v(n_{\square})=v}} \text{val}(\rho_2^v) \\
&= \sum_{v \in V_{\oplus}} f^v(t) g^v(c).
\end{aligned}$$

421 Let  $M_v \in K^{\text{Tree}(X) \times \text{Context}(X)}$  be the matrix given by  $M_v(t, c) =$   
422  $f^v(t) g^v(c)$ : its rank is at most one as  $M_v$  is the product of a column  
423 vector by a row vector. The previous equality reads in matrix form  
424  $H_f = \sum_{v \in V_{\oplus}} M_v$ . Hence, we obtain the announced lower bound  
425 using rank subadditivity:

$$\text{rank}(H_f) = \text{rank} \left( \sum_{v \in V_{\oplus}} M_v \right) \leq \sum_{v \in V_{\oplus}} \text{rank}(M_v) \leq |V_{\oplus}| = |\mathcal{C}|.$$

426 We now turn to the upper bound, and assume  $f$  is homoge-  
427 neous.

428 We first give a construction of a circuit, then provide and prove  
429 by induction a strong invariant which implies that the circuit does  
430 indeed compute  $f$ . For every  $t \in \text{Tree}(X)$ , we let  $H_t$  denote the  
431 corresponding column in the Hankel matrix, *i.e.*  $H_t : c \mapsto c[t]$ .

432 Let  $T \subseteq \text{Tree}(X)$  be such that  $(H_t)_{t \in T}$  is a basis of  $\{H_t \mid$   
433  $t \in \text{Tree}(X)\}$ . In particular  $T$  has size  $\text{rank}(H_f)$ . For any  $t' \in$   
434  $\text{Tree}(X)$ , we let  $\alpha_t^{t'}$  denote the coefficient of  $H_t$  in the decomposition  
435 of  $H_{t'}$  on  $(H_t)_{t \in T}$ , that is,

$$(\star) \quad H_{t'} = \sum_{t \in T} \alpha_t^{t'} H_t.$$

436 We may now explicitly define circuit  $\mathcal{C}$ :

- 437     ◦ The addition gates are (identified with) elements of  $T$ . The  
 438     output weight of  $t \in T$  is  $f(t)$ .
- 439     ◦ The input gates are given by elements of  $X$  (and the matching  
 440     label). The input gate  $x \in X$  has an outgoing arc to the  
 441     addition gate  $t \in T$  with weight  $\alpha_t^x$ .
- 442     ◦ The multiplication gates are given by elements  $(t_0, t_1, t) \in T^3$ .  
 443     Such a multiplication gate has an incoming arc from  $t_0$  on the  
 444     left, an incoming arc from  $t_1$  on the right, and an outgoing  
 445     arc to  $t$ , with weight  $\alpha_t^{t_1 t_2}$ .

446 Note that the size of  $\mathcal{C}$  is  $|T| = \text{rank}(H_f)$ .

447 For  $\mathcal{C}$  to be well-defined as a circuit, it remains to show that  
 448 its underlying graph is acyclic. This is implied by the fact that  
 449  $\alpha_t^{t_1 t_2}$  may only be non-zero if  $\deg(t) = \deg(t_1) + \deg(t_2)$ , which we  
 450 now prove. Since  $f$  is homogeneous of degree  $d$ ,  $H_t$  may be non-  
 451 zero only on contexts  $c$  such that  $\deg(c[t]) = d$ , that is,  $\deg(c) =$   
 452  $d - \deg(t) + 1$ . Hence, the set  $\{H_t, t \in T\}$  may be partitioned  
 453 according to the degree of  $t$  into parts with disjoint support, so  
 454 for the decomposition  $(\star)$  to hold, it must be that  $\alpha_t^{t'} \neq 0$  implies  
 455  $\deg(t) = \deg(t')$ .

456 For  $t \in T$ , we let  $g_t : \text{Tree}(X) \rightarrow K$  denote the polynomial  
 457 computed at gate  $t$  in  $\mathcal{C}$ . We will now show, by induction on the  
 458 size of  $t' \in \text{Tree}(X)$ , that  $g_t(t') = \alpha_t^{t'}$ .

459 If  $t' = x \in X$ , then  $g_t(t') = \alpha_t^x$ , so the base case is clear. We now  
 460 assume that  $t' = t'_1 \cdot t'_2 \in \text{Tree}(X)$ , and show that  $\sum_{t \in T} g_t(t')H_t =$   
 461  $H_{t'}$ , which is enough to conclude by uniqueness of the decomposi-  
 462 tion in  $(\star)$ . For that we will show that the previous equality holds  
 463 for any context  $c \in \text{Context}(X)$ .

464 We first remark the following

$$\begin{aligned}
 \sum_{t \in T} g_t(t') H_t &= \sum_{t \in T} \left( \sum_{t_1, t_2 \in T} \alpha_t^{t_1 \cdot t_2} g_{t_1}(t'_1) g_{t_2}(t'_2) \right) H_t \\
 &= \sum_{t \in T} \left( \sum_{t_1, t_2 \in T} \alpha_t^{t_1 \cdot t_2} \alpha_{t_1}^{t'_1} \alpha_{t_2}^{t'_2} \right) H_t \\
 &= \sum_{t_1, t_2 \in T} \alpha_{t_1}^{t'_1} \alpha_{t_2}^{t'_2} \left( \sum_{t \in T} \alpha_t^{t_1 \cdot t_2} H_t \right) \\
 &= \sum_{t_1, t_2 \in T} \alpha_{t_1}^{t'_1} \alpha_{t_2}^{t'_2} H_{t_1 \cdot t_2}.
 \end{aligned}$$

465 Now, let  $c \in \text{Context}(X)$ . For any tree  $t \in \text{Tree}(X)$ , we define  
 466  $c_t^1 = c[\square \cdot t] \in \text{Context}(X)$ , and  $c_t^2 = c[t \cdot \square] \in \text{Context}(X)$  (see  
 467 Figure 2.4). Then for any  $t_1, t_2, c[t_1 \cdot t_2] = c_{t_2}^1[t_1] = c_{t_1}^2[t_2]$ .

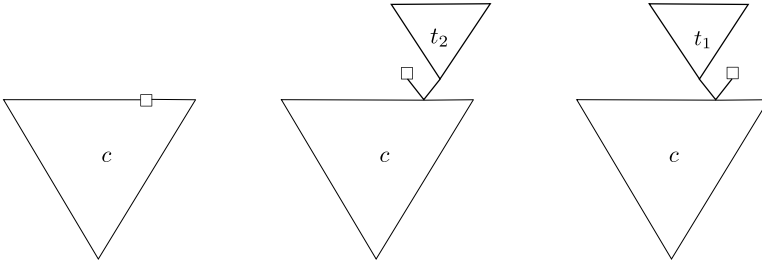


Figure 2.4: A context  $c$ , and the contexts  $c_{t_2}^1$  and  $c_{t_1}^2$ .

468 Evaluating at  $c$ , we now obtain

$$\begin{aligned}
 \sum_{t \in T} g_t(t') H_t(c) &= \sum_{t_1, t_2 \in T} \alpha_{t_1}^{t'_1} \alpha_{t_2}^{t'_2} H_{t_1 \cdot t_2}(c) = \sum_{t_1, t_2 \in T} \alpha_{t_1}^{t'_1} \alpha_{t_2}^{t'_2} f(c[t_1 \cdot t_2]) \\
 &= \sum_{t_1, t_2 \in T} \alpha_{t_1}^{t'_1} \alpha_{t_2}^{t'_2} f(c_{t_2}^1[t_1]) = \sum_{t_1, t_2 \in T} \alpha_{t_2}^{t'_2} H_{t_1}(c_{t_2}^1) \\
 &= \sum_{t_2 \in T} \alpha_{t_2}^{t'_2} H_{t'_1}(c_{t_2}^1) = \sum_{t_2 \in T} \alpha_{t_2}^{t'_2} H_{t'_1 \cdot t_2}(c) \\
 &= \sum_{t_2 \in T} \alpha_{t_2}^{t'_2} f(c_{t'_1}^2[t_2]) = \sum_{t_2 \in T} \alpha_{t_2}^{t'_2} H_{t_2}(c_{t'_1}^2) = H_{t_2}(c_{t'_1}^2) \\
 &= H_t(c)
 \end{aligned}$$

469 which proves the wanted invariant, namely  $g_t(t') = \alpha_t^{t'}$ . Hence, the  
 470 value computed by the circuit for monomial  $t'$  is precisely

$$\sum_{t \in T} g_t(t') f(t) = \sum_{t \in T} \alpha_t^{t'} H_t(\square) = H_{t'}(\square) = f(t'),$$

471 which concludes the proof of the upper bound. □

472 The remainder of this paper consists in applying Theorem 2.4 to  
 473 obtain lower bounds in various cases. To this end we need a better  
 474 understanding of the Hankel matrix: in Section 3 we introduce a  
 475 few concepts and in Section 4 we develop decomposition theorems  
 476 for the Hankel matrix.

477 Before digging any deeper we can already give two applications  
 478 of Theorem 2.4, yielding simple proofs of non-trivial results from  
 479 the literature.

480 The first lower bound we obtain is a separation of **VP** and  
 481 **VNP** in the commutative non-associative setting. It was already  
 482 obtained in (Hrubeš *et al.* 2010, Theorem 6).

483 Another early result is an alternative proof of (Arvind & Raja  
 484 2016, Theorem 26), which gives an exponential lower bound for  
 485 the permanent and the determinant against unambiguous circuits  
 486 in the *associative* setting.

### 487 Separation of Commutative Non-Associative VP and VNP.

488 We now give an alternative separation argument of the classes **VP**  
 489 and **VNP** in the commutative non-associative setting. The origi-  
 490 nal proof is due to (Hrubeš *et al.* 2010, Theorem 6), it exhibits  
 491 a polynomial which requires a superpolynomial circuit to be com-  
 492 puted. For simplicity, we give a slightly different polynomial, but  
 493 the proof is very much a reinterpretation of that of Hrubeš *et al.*  
 494 (2010) in the newly introduced vocabulary.

495 **COROLLARY 2.5.** *For  $d > 1$ , let  $f$  be the commutative non-associative*  
 496 *polynomial of degree  $2d$  and over two variables  $x_0$  and  $x_1$  defined*  
 497 *by*

$$f = \sum_{\varepsilon_1, \dots, \varepsilon_d \in \{0,1\}} (((\dots(x_{\varepsilon_1} x_{\varepsilon_2}) x_{\varepsilon_3}) \dots) x_{\varepsilon_d})^2.$$

498 *Any circuit computing  $f$  has size at least  $3 \times 2^{d-2}$ .*

499 PROOF. We give a lower bound on the rank of the Hankel matrix.  
 500 We consider the submatrix restricted to contexts with  $(d+1)$  leaves  
 501 of the form  $((\cdots(((x_{\varepsilon_1} \cdot x_{\varepsilon_2}) x_{\varepsilon_3}) x_{\varepsilon_4}) \cdots) x_{\varepsilon_d})\square)$  and to trees  
 502 with  $d$  leaves of the form  $((\cdots(((x_{\varepsilon'_1} \cdot x_{\varepsilon'_2}) x_{\varepsilon'_3}) x_{\varepsilon'_4}) \cdots) x_{\varepsilon'_d})$ . See  
 503 Figure 2.5 for a depiction.

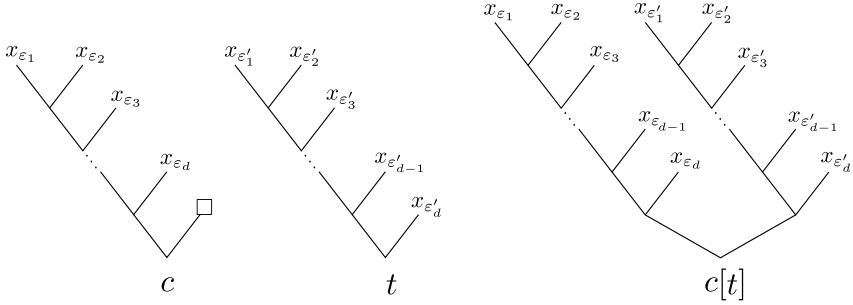


Figure 2.5: The context  $c = (((\cdots(((x_{\varepsilon_1} \cdot x_{\varepsilon_2}) x_{\varepsilon_3}) x_{\varepsilon_4}) \cdots) x_{\varepsilon_d})\square)$ , the tree  $t = ((\cdots(((x_{\varepsilon'_1} \cdot x_{\varepsilon'_2}) x_{\varepsilon'_3}) x_{\varepsilon'_4}) \cdots) x_{\varepsilon'_d})$  and their composition  $c[t]$ .

504 This matrix is a permutation matrix of size  $3 \times 2^{d-2}$ , which is,  
 505 up to commutativity, the number of different trees or contexts of  
 506 the form mentioned above.  $\square$

507 We now present a first lower bound in the *associative* setting.  
 508 The method we shall use is generic: consider an associative circuit  
 509  $\mathcal{C}$ , from a given restricted class of circuits, computing a given poly-  
 510 nomial  $f$ . Let  $\tilde{f}$  be the non-associative polynomial computed by  
 511  $\mathcal{C}$  when it is seen as non-associative. The restriction on  $\mathcal{C}$  together  
 512 with the coefficients in  $f$  provide informations on  $\tilde{f}$  which we use  
 513 to derive a lower bound on  $\text{rank}(H)$ , which is also a lower bound  
 514 on  $\mathcal{C}$  thanks to Theorem 2.4.

515 **Lower Bound Against Associative Unambiguous Circuits.**

516 We give a lower bound for unambiguous circuits computing the  
 517 *associative* permanent or determinant. A circuit is said **unam-**  
 518 **biguous**, if for each (associative) monomial  $m$ , there is at most  
 519 one tree  $t$  labelled by  $m$  such that  $\mathcal{C}$  has a run over  $t$ . Such cir-  
 520 cuits were already studied in Arvind & Raja (2016), in which the  
 521 authors provide a lower bound for the permanent: we show how to

522 recover their result using the Hankel matrix. Note that this notion  
 523 makes sense in both the commutative and the non-commutative  
 524 settings and that our lower bounds hold in both settings.

Recall that, on variables  $X = \{x_{i,j} \mid i, j \in [n]\}$ , if one lets  $S_n$  denote the set of all permutations over  $[n]$  and  $\text{sgn}(\sigma)$  denote the signature of a permutation  $\sigma$ , the determinant of degree  $n$  is the polynomial

$$\text{Det} = \sum_{\sigma \in S_n} \prod_{i=1}^n \text{sgn}(\sigma) x_{i,\sigma(i)}$$

and the permanent of degree  $n$  is the polynomial

$$\text{Per} = \sum_{\sigma \in S_n} \prod_{i=1}^n x_{i,\sigma(i)}.$$

525

526 **COROLLARY 2.6.** *Any unambiguous circuit computing the deter-*  
 527 *minant or the permanent of degree  $n$  has size at least  $\binom{n}{n/3}$ .*

528 **PROOF.** Consider an unambiguous circuit  $\mathcal{C}$  computing the per-  
 529 manent (the proof is easily adapted to a circuit computing the de-  
 530 terminant) of degree  $n$  on variables  $X = \{x_{i,j} \mid i, j \in [n]\}$ . For any  
 531 permutation  $\sigma$ , let  $t_\sigma \in \text{Tree}(X)$  be the unique (non-associative)  
 532 monomial along which there is a run computing the (associative)  
 533 monomial  $x_{1,\sigma(1)}x_{2,\sigma(2)} \cdots x_{n,\sigma(n)}$ . Then, the non-associative poly-  
 534 nomial  $\tilde{f}$  computed by  $\mathcal{C}$  when it is seen as a non-associative circuit  
 535 is precisely  $\tilde{f} = \sum_{\sigma} t_\sigma$ . According to Theorem 2.4, it suffices to  
 536 lower bound the rank of  $H_{\tilde{f}}$ .

Let  $(A, S) \subseteq [n]^2$  be a pair of subsets. We let  $T_{A \rightarrow S} \subseteq \text{Tree}(X)$  be the subset of trees  $t$  such that the set of first (*resp.* second) indices of the labels of  $t$  is precisely  $A$  (*resp.*  $S$ ). Symmetrically, let  $C_{A \rightarrow S} \subseteq \text{Context}(X)$  be the subset of contexts  $c$  such that the set of first (*resp.* second) indices of the labels (except for the  $\square$ ) of  $c$  is precisely  $[n] \setminus A$  (*resp.*  $[n] \setminus S$ ). If  $(A, S) \neq (A', S')$ , then  $T_{A \rightarrow S}$  and  $T_{A' \rightarrow S'}$  are disjoint, as is the case for  $C_{A \rightarrow S}$  and  $C_{A' \rightarrow S'}$ . Moreover, if  $t \in T_{A \rightarrow S}$  and  $c \in C_{A' \rightarrow S'}$ , it must be that  $\tilde{f}(c[t]) = 0$ . Hence,  $H_{\tilde{f}}$  is a block-diagonal matrix, with blocks  $H_{A,S}$  being given by

restricting the columns to some  $T_{A \rightarrow S}$  and the rows to  $C_{A \rightarrow S}$ . Note that if  $|A| \neq |S|$  then  $H_{A,S} = 0$ . In particular,

$$\text{rank}(H_{\tilde{f}}) = \sum_{\substack{A, S \subseteq [n] \\ |A|=|S|}} \text{rank}(H_{A,S}).$$

537 We now show using a counting argument that an exponential  
538 number of such blocks are non-zero and hence, have rank at least 1.

539 For all permutations  $\sigma$ , we choose a subtree  $t'_\sigma$  of  $t_\sigma$  which has  
540 size in  $[n/3, 2n/3]$ , and let  $(A_\sigma, S_\sigma)$  be such that  $t'_\sigma \in T_{A_\sigma \rightarrow S_\sigma}$ .  
541 Note that  $n/3 \leq |A_\sigma| = |S_\sigma| = |t'_\sigma| \leq 2n/3$ , and that  $H_{A_\sigma, S_\sigma} \neq 0$ .  
542 Moreover, it must be that  $\sigma(A_\sigma) = S_\sigma$ . Hence, if  $A, S \subseteq [n]$  are  
543 fixed such that  $n/3 \leq |A| = |S| \leq 2n/3$ ,

$$|\{\sigma \mid A_\sigma = A \text{ and } S_\sigma = S\}| \leq |\{\sigma \mid \sigma(A) = S\}| \leq \left(\frac{n}{3}\right)! \left(\frac{2n}{3}\right)!$$

544 Hence, the number of non-zero blocks  $H_{A,S}$  is at least

$$\frac{n!}{\left(\frac{n}{3}\right)! \left(\frac{2n}{3}\right)!} = \binom{n}{n/3}$$

545 which concludes the proof.  $\square$

546 Note that this exact proof goes beyond the case of unambiguous  
547 circuits. It is actually sufficient to assume that all non-associative  
548 monomials  $t$  such that  $\tilde{f}(t) \neq 0$  are labelled by a monomial of the  
549 form  $x_{1,\sigma(1)}x_{2,\sigma(2)} \cdots x_{n,\sigma(n)}$  for some permutation  $\sigma$ .

### 550 3. Decomposing the Hankel Matrix: Unique 551 Parse Tree Circuits

552 Theorem 2.4, as already illustrated by Corollary 2.6, is a natural  
553 tool to derive lower bounds thanks to an analysis of the rank of  
554 the Hankel Matrix. In order to lower bound this rank for the  
555 most general classes possible, we need tools, parse trees and partial  
556 derivative matrices, that we introduce now; we then apply these  
557 tools to derive a general result regarding the class of Unique Parse  
558 Tree circuits (Theorem 3.9). In Section 4, we will push this analysis  
559 further and derive generic lower bounds.

560 **3.1. Parse Trees.** With any monomial  $t \in \text{Tree}(X)$  we associate  
 561 its *shape*  $\text{shape}(t) \in \text{Tree}$  as the tree obtained from  $t$  by removing  
 562 the labels at the leaves.

563 **DEFINITION 3.1.** *Let  $\mathcal{C}$  be a circuit computing a non-commutative*  
 564 *non-associative polynomial  $f$ . A **parse tree** of  $\mathcal{C}$  is any shape*  
 565  *$s \in \text{Tree}$  for which there exists a monomial  $t \in \text{Tree}(X)$  whose*  
 566 *coefficient in  $f$  is non-zero and such that  $s = \text{shape}(t)$ . We let*  
 567  *$PT(\mathcal{C}) = \{\text{shape}(t) \mid f(t) \text{ non-zero}\}$ .*

568 The notion of parse trees has been considered in many previous  
 569 works, see for example (Jerrum & Snir 1982; Allender *et al.* 1998;  
 570 Malod & Portier 2008; Arvind & Raja 2016; Lagarde *et al.* 2016;  
 571 Saptharishi & Tengse 2017; Lagarde *et al.* 2018).

572 **REMARK 3.2.** *Let  $\mathcal{C}$  be a circuit computing a homogeneous poly-*  
 573 *nomial of degree  $d$ . Then asymptotically,  $|PT(\mathcal{C})| \leq 4^d$ . Indeed,*  
 574 *the maximal number of parse trees corresponds to the number of*  
 575 *ordered binary trees with  $d$  leaves which is the  $(d - 1)$ -th Catalan*  
 576 *number  $C_{d-1}$ . Asymptotically, one has  $C_k \sim \frac{4^k}{k^{3/2}\sqrt{\pi}}$  which implies*  
 577 *the announced lower bound on the number of parse trees.*

578 **3.2. Partial Derivative Matrices.** We now introduce a popu-  
 579 lar tool for proving circuit lower bounds, namely, partial derivative  
 580 matrices, originated from (Hyafil 1977; Nisan 1991) and widely  
 581 used and extended in subsequent works, see for example (Nisan &  
 582 Wigderson 1997; Dvir *et al.* 2012; Gupta *et al.* 2014; Kayal *et al.*  
 583 2014a; Limaye *et al.* 2016; Kumar & Saraf 2017).

584 For  $A \subseteq [d]$  of size  $i$ ,  $u \in X^{d-i}$ , and  $v \in X^i$ , we define the  
 585 monomial  $u \otimes_A v \in X^d$ : it is obtained by interleaving  $u$  and  $v$   
 586 with  $u$  taking the positions indexed by  $[d] \setminus A$  and  $v$  the positions  
 587 indexed by  $A$ . For instance  $x_1x_2 \otimes_{\{2,4\}} y_1y_2 = x_1y_1x_2y_2$ .

588 **DEFINITION 3.3.** *Let  $f$  be a homogeneous non-commutative as-*  
 589 *sociative polynomial. Let  $A \subseteq [d]$  be a set of positions of size*  
 590  *$i$ .*

591 *The **partial derivative matrix**  $M_A(f)$  of  $f$  with respect to*  
 592  *$A$  is defined as follows: the rows are indexed by  $u \in X^{d-i}$  and the*



593 columns by  $v \in X^i$ , and the value of  $M_A(f)(u, v)$  is the coefficient  
 594 of the monomial  $u \otimes_A v$  in  $f$ .

595 **REMARK 3.4.** *The terminology partial derivative matrix, widely*  
 596 *adopted in the literature, comes from the observation that the row*  
 597 *labelled by monomial  $u$  of the matrix contains the coefficients of the*  
 598 *partial derivative  $\frac{\partial f}{\partial u}$ . The same remark can be made for columns*  
 599 *of  $M_A(f)$ . This will not be exploited in this paper.*

600 **EXAMPLE 3.5.** Let  $f = xyxy + 3xxyy + 2xxxy + 5yyyy$  and  $A =$   
 601  $\{2, 4\}$ . Then  $M_A(f)$  is given below.

	$xx$	$xy$	$yx$	$yy$
$xx$	0	2	0	1
$yx$	0	0	0	0
$xy$	0	3	0	0
$yy$	0	0	0	5

602

◇

We define a distance  $\text{dist} : \mathcal{P}([d]) \times \mathcal{P}([d]) \rightarrow \mathbb{N}$  on subsets of  $[d]$  by letting  $\text{dist}(A, B)$  be the minimal number of additions and deletions of elements of  $[d]$  to go from  $A$  to  $B$ , assuming that complementing is for free. Formally,

$$\text{dist}(A, B) = \min\{|\Delta(A, B)|, |\Delta(A^c, B)|\},$$

603 where  $\Delta(A, B) = (A \setminus B) \cup (B \setminus A)$  is the symmetric difference  
 604 between  $A$  and  $B$ . This is illustrated in Figure 3.1.

605 **REMARK 3.6.** *A similar looking notion of distance is also available*  
 606 *in the current literature for commutative depth-4 lower bounds.*  
 607 *This was first implicitly defined by Fournier et al. (2014) and*  
 608 *by Kayal et al. (2014a), and later made explicit by Chillara &*  
 609 *Mukhopadhyay (2019).*

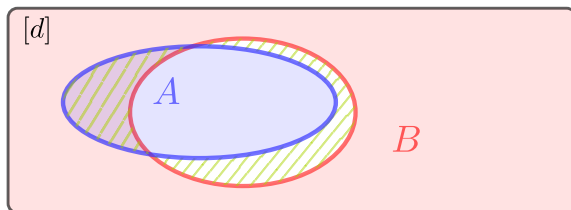


Figure 3.1: In this case, the symmetric difference is smaller when complementing one of the sets, so  $\text{dist}(A, B)$  is the cardinality of the hatched subset.

610 REMARK 3.7. *The apparent asymmetry in the definition is artificial as it does hold that  $\text{dist}(A, B) = \text{dist}(B, A)$ . It is also the case*  
 611 *that  $\text{dist}(A, B) = 0 \implies A = B$  or  $A = B^c$ . In fact  $\text{dist}$  is indeed*  
 612 *a distance over subsets of  $[d]$  modulo complementation.*  
 613

614 The following lemma (see e.g., (Limaye *et al.* 2016)) informally  
 615 says that, if  $A$  and  $B$  are close to each other, then the ranks of the  
 616 corresponding partial derivative matrices are close to each other as  
 617 well. Though it is well known, we give a proof for completeness.

618 LEMMA 3.8. *Let  $f$  be a homogeneous non-commutative associa-*  
 619 *tive polynomial of degree  $d$  with  $n$  variables. Then, for any subsets*  
 620  *$A, B \subseteq [d]$ ,  $\text{rank}(M_A(f)) \leq n^{\text{dist}(A,B)} \text{rank}(M_B(f))$ .*

621 PROOF. Without loss of generality, one may safely assume that  
 622  $\text{dist}(A, B) = |\Delta(A, B)|$  (by transposing the matrix  $M_A(f)$  if nec-  
 623 essary).

624 We prove the statement by induction on  $d = |\Delta(A, B)|$ . If  
 625  $d = 0$ , this is trivial since  $A$  and  $B$  are identical in this case. For  
 626 the case  $d = 1$ , let us assume that  $A = B \cup \{i\}$  (the other case being  
 627 very similar). We divide  $M_A(f)$  into horizontal blocks, one for each  
 628 variable  $x$ , that we call  $M_A(f)^x$ , corresponding to the monomials  
 629 for which the position  $i$  is occupied by the variable  $x$ . Therefore  
 630 the rank of  $M_A(f)$  is upper bounded by  $\sum_x \text{rank}(M_A(f)^x)$ , but  
 631 each  $M_A(f)^x$  is a submatrix of  $M_B(f)$  so that  $\text{rank}(M_A(f)^x) \leq$   
 632  $\text{rank}(M_B(f))$ , hence the result.

633 If  $d > 1$ , we first find a set  $C$  such that  $|\Delta(A, C)| = 1$  and

634  $|\Delta(C, B)| = d - 1$ , and we conclude by applying the induction  
 635 hypothesis and using the case  $d = 1$ .  $\square$

636 At this point, we have the material in hands to describe a pre-  
 637 cise characterization of the size of the smallest Unique Parse Tree  
 638 circuit which computes a given polynomial. We take this short de-  
 639 tour before moving on to our core lower bound results in Section 4.

640 **3.3. Characterization of Smallest Unique Parse Tree Cir-**  
 641 **cuit.** Unique Parse Tree (UPT) circuits are non-commutative as-  
 642 sociative circuits with a unique parse tree. They were first intro-  
 643 duced in (Lagarde *et al.* 2016). They generalize ABPs, which are  
 644 equivalent to UPT circuits with a left comb as their unique parse  
 645 tree (a left comb being a tree corresponding to the shape of tree  $t$   
 646 in Figure 2.5). Hence, we recover Nissan’s Theorem (Nisan 1991)  
 647 when instantiating our characterization result, Theorem 3.9, to left  
 648 combs. Our techniques allow a slight improvement and a better  
 649 understanding of their results since the original result requires a  
 650 normal form which can lead to an exponential blow-up.

651 Given a shape  $s \in \text{Tree}$  of size  $d$ , i.e., with  $d$  leaves and a node  
 652  $v$  of  $s$ , we let  $s_v$  denote the subtree of  $s$  rooted in  $v$ , and  $I_v \subseteq [d]$   
 653 denote the interval of positions of the leaves of  $s_v$  in  $s$ . We say  
 654 that  $s' \in \text{Tree}$  is a **subshape** of  $s$  if  $s' = s_v$  for some  $v$ , and that  
 655  $I \subseteq [d]$  is spanned by  $s$  if  $I = I_v$  for some  $v$ . Figure 3.2 illustrates  
 656 the occurrences of a subshape in a shape.

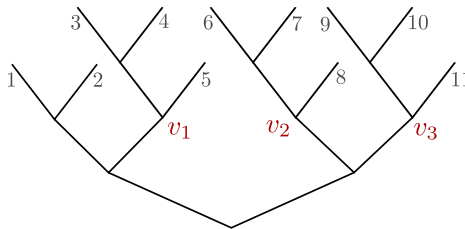


Figure 3.2: A shape of size 11, in which three nodes  $v_1, v_2, v_3$  span the same subshape. The corresponding spanned intervals are  $I_{v_1} = [3, 5], I_{v_2} = [6, 8]$  and  $I_{v_3} = [9, 11]$ . We display the position of each leaf for readability.

657 Let  $f$  be a homogeneous non-commutative associative polyno-

658 mial of degree  $d$ , let  $s \in \text{Tree}$  be a shape of size  $d$ , and let  $s'$  be  
 659 a subshape of  $s$  such that  $v_1, \dots, v_p$  are all the nodes  $v$  of  $s$  such  
 660 that  $s' = s_v$ . We define

$$M_{s'} = \begin{bmatrix} M_{I_{v_1}}(f) \\ M_{I_{v_2}}(f) \\ \vdots \\ M_{I_{v_p}}(f) \end{bmatrix}.$$

661 THEOREM 3.9. *Let  $f$  be a homogeneous non-commutative asso-*  
 662 *ciative polynomial of degree  $d$  and let  $s \in \text{Tree}$  be a shape of size*  
 663  *$d$ . Then the smallest UPT circuit with shape  $s$  computing  $f$  has*  
 664 *size exactly*

$$\sum_{s' \text{ subshape of } s} \text{rank}(M_{s'}).$$

665 PROOF. Let  $\mathcal{C}$  be a UPT circuit with shape  $s$  computing  $f$ . We  
 666 let  $\tilde{f}$  denote the non-associative polynomial computed by  $\mathcal{C}$ .

667 Since  $\mathcal{C}$  is UPT with shape  $s$ ,  $\tilde{f}$  is the *unique* non-associative  
 668 polynomial which is non-zero only on trees with shape  $s$  and projects  
 669 to  $f$ , i.e.,  $\tilde{f}(t) = f(u)$  if  $\text{shape}(t) = s$  and  $t$  is labelled by  $u$ , and  
 670  $\tilde{f}(t) = 0$  otherwise.

671 In particular, the size of the smallest UPT circuit with shape  
 672  $s$  computing  $f$  is the same as the size of the smallest circuit com-  
 673 puting  $\tilde{f}$ , which thanks to Theorem 2.4 is equal to the rank of the  
 674 Hankel matrix  $H_{\tilde{f}}$ .

675 The Hankel matrix of  $\tilde{f}$  may be non-zero only on columns in-  
 676 dexed by trees whose shapes  $s'$  are subshapes of  $s$ , and on such  
 677 columns, non-zero values are on rows corresponding to a context  
 678 obtained from  $s$  by replacing an occurrence of  $s'$  by  $\square$ . The corre-  
 679 sponding blocks are precisely the matrices  $M_{s'}$ , and are placed in  
 680 a diagonal fashion, hence the lower bound.  $\square$

681 Theorem 3.9 can be applied to concrete polynomials, for in-  
 682 stance to the permanent of degree  $d$ .

683 COROLLARY 3.10. *Let  $s \in \text{Tree}$  be a shape. The smallest UPT*  
 684 *circuit with shape  $s$  computing the permanent has size*

$$\sum_{v \text{ node of } s} \binom{d}{|I_v|},$$

685 *where  $I_v$  is the set of leaves in the subtree rooted at  $v$  in  $s$ . In*  
 686 *particular, this is always larger than  $\binom{d}{d/3}$ .*

PROOF. Let  $s'$  be a subshape of  $s$ , and  $v_1, \dots, v_p$  be all the nodes of  $s$  such that  $s_{v_i} = s'$ . Let  $\ell = |I_{v_i}|$  which does not depend on  $i$ . There are no  $i \neq j$  such that  $v_i$  is a descendant of  $v_j$ , so the  $I_{v_i}$  are pairwise disjoint. Let  $I_{v_i} = [a_i, a_i + \ell - 1]$ . The coefficient of  $M_{I_{v_i}}(\text{Per})$  in  $(u, w) \in X^{d-\ell} \times X^\ell$ , namely,  $\text{Per}(u \otimes_{I_{v_i}} w)$ , may be non-zero only if  $w$  is of the form

$$x_{a_i, b_1} x_{a_i+1, b_2} \cdots x_{a_i+\ell-1, b_\ell}$$

for some  $b_1, \dots, b_\ell \in [d]$ . In particular, the  $M_{I_{v_i}}(\text{Per})$  have non-zero columns with disjoint supports, so

$$\text{rank}(M_{s'}) = \sum_i \text{rank}(M_{I_{v_i}}(\text{Per})).$$

687 We claim now that  $\text{rank}(M_{I_{v_i}}(\text{Per})) = \binom{d}{\ell}$ , which leads to the  
 688 announced formula. Indeed, any subset  $A$  of  $[d]$  of size  $\ell$  corre-  
 689 sponds to a block full of 1's in the matrix  $M_{I_{v_i}}(\text{Per})$  in the follow-  
 690 ing way:  $\text{Per}(u \otimes_{I_{v_i}} w) = 1$  whenever  $u$  is a monomial whose first  
 691 indices are  $[d] \setminus I_{v_i}$  and the second indices are any permutation of  
 692  $[d] \setminus A$ , and  $w$  is a monomial whose first indices are  $I_{v_i}$  and the  
 693 second indices are any permutation of  $A$ . Two such blocks have  
 694 disjoint rows and columns, and these are the only 1's in  $M_{I_{v_i}}(\text{Per})$ .  
 695 Moreover, there are  $\binom{d}{\ell}$  such sets  $A$ .

□

696 Applied to  $s$  being a left-comb, Corollary 3.10 yields that the  
 697 smallest ABP computing the permanent has size  $2^d + d$ . Applied  
 698 to  $s$  being a complete binary tree of depth  $k = \log d$ , the size of  
 699 the smallest UPT is  $\Theta\left(\frac{2^d}{d}\right)$ .

## 4. Decomposing the Hankel Matrix: Generic Lower Bounds

700  
701

702 We now get to the technical core of the paper where we establish  
703 generic lower bound theorems through a decomposition of the Han-  
704 kel matrix, that we will later instantiate in Section 5 to concrete  
705 classes of circuits.

706 We first restrict ourselves to the non-commutative setting. Our  
707 first decomposition, Theorem 4.1, seems to capture mostly pre-  
708 viously known techniques. However, the second, more powerful,  
709 decomposition, Theorem 4.2, takes advantage of the global shape  
710 of the Hankel matrix. Doing so allows to go beyond previous re-  
711 sults only hinging around considering partial derivatives matrices  
712 which turn out to be isolate slices of the Hankel matrix.

713 We later explain in Section 4.3 how to extend the study to the  
714 commutative case.

715 **4.1. General Roadmap.** Let  $f$  be a (commutative or non-com-  
716 mutative) associative polynomial for which we want to prove lower  
717 bounds. Consider a circuit  $\mathcal{C}$  which computes  $f$ , and let  $\tilde{f}$  be the  
718 non-associative polynomial computed by  $\mathcal{C}$ . Our aim is, following  
719 Theorem 2.4, to lower bound the rank of the Hankel matrix  $H_{\tilde{f}}$ .  
720 We know that polynomials  $\tilde{f}$  and  $f$  are equal up to associativity,  
721 which provides linear relations among the coefficients of  $H_{\tilde{f}}$ .

722 The bulk of the technical work is to reorganize the rows and  
723 columns of  $H_{\tilde{f}}$  in order to decompose it into blocks which may  
724 be identified as partial derivative matrices with respect to some  
725 subsets  $A_1, A_2, \dots \subseteq [d]$ , of some associative polynomials which  
726 depend on  $\tilde{f}$  and sum to  $f$ . The number and choice of these subsets  
727 depend on the parse trees of the circuit  $\mathcal{C}$ .

728 Now, assume that there exists a subset  $A \subseteq [d]$  which is at  
729 distance at most  $\delta$  to each  $A_i$ . Losing a factor of  $n^\delta$  on the rank  
730 through the use of Lemma 3.8 we reduce the aforementioned blocks  
731 of  $H_{\tilde{f}}$  to partial derivatives with respect to  $A$ . Such matrices can  
732 then be summed to recover the partial derivative matrix of  $f$  with  
733 respect to  $A$ , yielding in the lower bound a (dominating) factor of  
734  $\text{rank}(M_A(f))$ .

735 **4.2. Generic Lower Bounds in the Non-commutative Set-**  
 736 **ting.** Following the general roadmap described above, we obtain  
 737 a first generic lower bound result.

738 **THEOREM 4.1.** *Let  $f$  be a non-commutative homogeneous poly-*  
 739 *nomial of degree  $d$  computed by a circuit  $\mathcal{C}$ . Let  $A \subseteq [d]$  and  $\delta \in \mathbb{N}$*   
 740 *such that all parse trees of  $\mathcal{C}$  span an interval at distance at most*  
 741  *$\delta$  from  $A$ . Then  $\mathcal{C}$  has size at least  $\text{rank}(M_A(f)) n^{-\delta} |\text{PT}(\mathcal{C})|^{-1}$ .*

742 **PROOF.** The proof relies on a better understanding of the struc-  
 743 ture of the Hankel matrix  $H = H_{\tilde{f}}$  of a general non-associative  
 744 polynomial  $\tilde{f} : \text{Tree}(X) \rightarrow K$ .

745 More precisely, we organize the columns and rows of  $H$  in or-  
 746 der to write it as a block matrix in which we can identify and  
 747 understand the blocks in terms of partial derivative matrices of  
 748 some non-commutative (but associative) polynomials which will  
 749 eventually correspond to parse trees. In the following we refer to  
 750 Figure 4.1 for illustration of the decompositions.

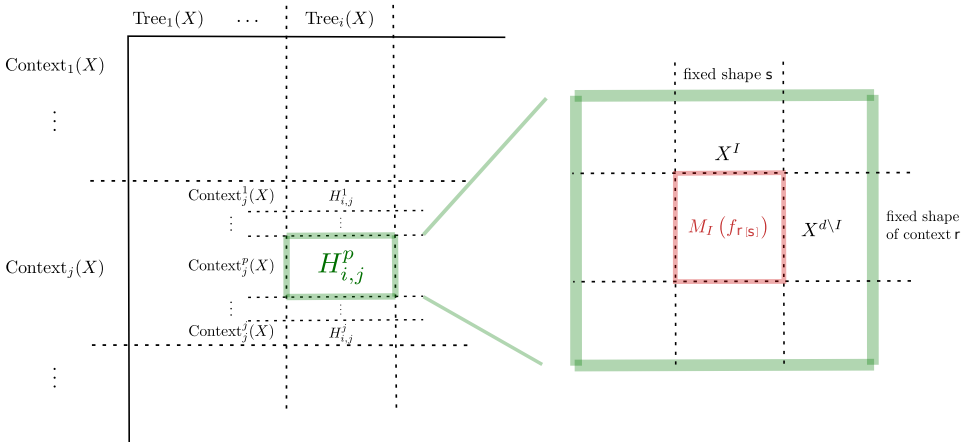


Figure 4.1: Decomposing  $H$  as blocks  $H_{i,j}^p$ , which further decompose into partial derivative matrices. Here,  $I$  denotes the interval  $[p, p + i - 1]$ .

751 Recall that  $\text{Tree}_k(X) \subseteq \text{Tree}(X)$  denotes the set of trees with  
 752  $k$  leaves, and let  $\text{Context}_k(X) \subseteq \text{Context}(X)$  denote the set of

753 contexts with  $k$  leaves (among which one is labelled by  $\square$ ). Note  
 754 that any tree  $t \in \text{Tree}_d(X)$  decomposes into  $2d - 1$  different pairs  
 755  $(t', c) \in \text{Tree}_k(X) \times \text{Context}_{d-k+1}(X)$  for some  $k$ , such that  $c[t'] = t$ ,  
 756 which correspond to the  $2d - 1$  nodes in  $t$ . We further partition  
 757  $\text{Context}_k(X) = \bigcup_{p=1}^k \text{Context}_k^p(X)$ , with  $\text{Context}_k^p(X)$  being the  
 758 set of contexts where  $\square$  is on the  $p$ -th leaf.

759 Using these partitions for trees and contexts, we may write  $H$   
 760 as a block matrix with blocks  $H_{i,j} = H_{|\text{Tree}_i(X) \times \text{Context}_j(X)}$ . Using the  
 761 finer refinement of contexts, we write block  $H_{i,j}$  as a tower (recall  
 762 that contexts label the rows of  $H$ ) of sub-blocks  $H_{i,j}^p$ , for  $p \in [j]$ ,  
 763 where  $H_{i,j}^p = H_{|\text{Tree}_i(X) \times \text{Context}_j^p(X)}$ . We now focus on  $H_{i,j}^p$ , which  
 764 we will further decompose into blocks that are partial derivative  
 765 matrices of some homogeneous non-commutative polynomials on  
 766 the interval  $[p, p + i - 1]$ .

As  $\text{Tree}_i(X)$  is the set of trees with  $i$  leaves, it can be seen  
 as all possible labeling of shapes with  $i$  leaves by variables in  $X$ .  
 Hence,  $\text{Tree}_i(X) \simeq \text{Tree}_i \times X^i \simeq \text{Tree}_i \times X^{[p, p+i-1]}$ . Likewise,  
 $\text{Context}_j^p(X)$  is the set of contexts with  $j$  leaves and  $\square$  on the  $p$ -th  
 leaf, which can be seen as  $\text{Context}_j^p(X) \simeq \text{Context}_j^p \times X^{j-1} \simeq$   
 $\text{Context}_j^p \times X^{[1, i+j-1] \setminus [p, p+i-1]}$ , where  $\text{Context}_j^p$  is the set of contexts  
 of size  $j$  with no labels, except for a unique  $\square$  on the  $p$ -th leaf.  
 We now let, for any shape  $s \in \text{Tree}_{i+j-1}$ , the non-commutative  
 (but associative) homogeneous polynomial  $f_s$  of degree  $i + j - 1$  be  
 defined by

$$f_s : X^{i+j-1} \rightarrow K$$

$$u \mapsto \tilde{f}(s \text{ labelled by } u)$$

767 Now, grouping the columns  $t \in \text{Tree}_i(X)$  of  $H_{i,j}^p$  which corre-  
 768 spond to the same shape  $s \in \text{Tree}_i$ , and the rows  $c \in \text{Context}_j^p(X)$   
 769 which correspond to the same shape (of context)  $r \in \text{Context}_j^p$ ,  
 770 we obtain a block matrix, in which the block indexed by  $(s, r)$  is  
 771 precisely the partial derivative matrix  $M_{[p, p+i-1]}(f_{r[s]})$ .

772 In the following, we will be interested in non-associative poly-  
 773 nomials  $\tilde{f} : \text{Tree}(X) \rightarrow K$  which project to a given associative



774  $f : X^* \rightarrow K$ , meaning that for each  $u \in X^*$ ,

$$\sum_{\substack{t \in \text{Tree}(X) \\ \text{label}(t)=u}} \tilde{f}(t) = f(u).$$

775 In this setting, one can see the decomposition  $f = \sum_{s \in \text{Tree}} f_s$  as a  
 776 decomposition over parse trees of a circuit computing  $f$ ,  $f_s$  being  
 777 the contribution of the parse tree  $s$  in the computation of  $f$ . We  
 778 have seen that if  $I = [p, p + i - 1]$  is an interval such that  $s$  de-  
 779 composes into  $s = r[s']$  for  $(s', r) \in \text{Tree}_i \times \text{Context}_j^p$ , which means  
 780 that  $I$  is spanned by  $s$ , then  $M_I(f_s)$  appears as a sub-matrix of  $H$ .  
 781 Hence,

$$(\star) \quad \max_{\substack{s \in \text{Tree} \\ I \text{ spanned by } s}} \text{rank}(M_I(f_s)) \leq \text{rank}(H).$$

782 Now, we have all the necessary tools to prove Theorem 4.1.  
 783 Let  $\tilde{f} : \text{Tree}(X) \rightarrow K$  be the non-associative polynomial computed  
 784 by  $\mathcal{C}$  when it is seen as a non-associative circuit. For any shape  
 785  $s \in \text{Tree}_d$ , let  $f_s : X^d \rightarrow K$  be defined as previously. In particular,  
 786  $\sum_{s \in \text{PT}(\mathcal{C})} f_s = f$ .

787 With a shape  $s \in \text{PT}(\mathcal{C})$ , we associate an interval  $I(s)$  spanned  
 788 by  $s$  and such that  $\text{dist}(A, I(s)) \leq \delta$ . Then we have

$$\begin{aligned} \text{rank}(M_A(f)) &= \text{rank} \left( \sum_{s \in \text{PT}(\mathcal{C})} M_A(f_s) \right) \\ &\leq \sum_{s \in \text{PT}(\mathcal{C})} \text{rank}(M_A(f_s)) && \text{by rank subadditivity} \\ &\leq \sum_{s \in \text{PT}(\mathcal{C})} n^\delta \text{rank}(M_{I(s)}(f_s)) && \text{by Lemma 3.8} \\ &\leq |\text{PT}(\mathcal{C})| n^\delta \text{rank}(H) && \text{by equation } (\star) \end{aligned}$$

789 Since, by Theorem 2.4,  $\text{rank}(H) \geq \text{rank}(M_A(f)) n^{-\delta} |\text{PT}(\mathcal{C})|^{-1}$   
 790 is a lower bound on  $|\mathcal{C}|$ , we obtain the announced result.  $\square$

791 The crux to prove Theorem 4.1 is to identify for each parse  
 792 tree  $s$  of  $\mathcal{C}$  a block in  $H_{\tilde{f}}$  containing the partial derivative matrix

793  $M_{I(s)}(f_s)$  where  $f_s$  is the polynomial corresponding to the contri-  
 794 bution of the parse tree  $s$  in the computation of  $f$  and  $I(s)$  is an  
 795 interval spanned by  $s$ .

796 However, we do not consider in this analysis how these blocks  
 797 are located relative to each other. A more careful analysis of  $H_{\tilde{f}}$   
 798 consists in grouping together all parse trees that lead to the same  
 799 spanned interval. Aligning and then summing these blocks we  
 800 replace the dependence in  $|\text{PT}(\mathcal{C})|$  by  $d^2$  which corresponds to  
 801 the total number of possibly spanned intervals of  $[d]$ . This yields  
 802 Theorem 4.2.

803 **THEOREM 4.2.** *Let  $f$  be a non-commutative homogeneous poly-*  
 804 *nomial of degree  $d$  computed by a circuit  $\mathcal{C}$ . Let  $A \subseteq [d]$  and  $\delta \in \mathbb{N}$*   
 805 *such that all parse trees of  $\mathcal{C}$  span an interval at distance at most*  
 806  *$\delta$  from  $A$ . Then  $\mathcal{C}$  has size at least  $\text{rank}(M_A(f)) n^{-\delta} d^{-2}$ .*

807 **REMARK 4.3.** *Note that this is an important improvement since*  
 808 *the number of parse trees can be up to about  $4^d$  (as noticed in Re-*  
 809 *mark 3.2). As we shall see in Section 5 the lower bounds we obtain*  
 810 *using Theorem 4.1 match known results, while using Theorem 4.2*  
 811 *yields substantial improvements.*

812 Before going on to the formal proof of Theorem 4.2, we start  
 813 by giving a high-level interpretation of the techniques used to go  
 814 from Theorem 4.1 to Theorem 4.2. Our aim is still to lower bound  
 815 the rank of the Hankel matrix  $H = H_{\tilde{f}}$  of some (unknown) non-  
 816 associative polynomial  $\tilde{f}$ , under the constraints that, for each  $u \in$   
 817  $X^*$ ,

$$\sum_{\substack{t \in \text{Tree}(X) \\ \text{label}(t)=u}} \tilde{f}(t) = f(u),$$

818 for some non-commutative (but associative) polynomial  $f : X^* \rightarrow$   
 819  $K$  that we control. Given the form of our constraints, a natural  
 820 strategy would be to sum some well chosen sub-matrices of  $H$  in  
 821 order to obtain a matrix that depends only on  $f$ , which we could  
 822 choose to have high rank.

823 As exposed earlier when proving Theorem 4.1, it is possible  
 824 to decompose  $f$  as the sum of some  $f_s$ 's, where  $s$  ranges over the

825 shapes used by  $\tilde{f}$ , and then obtain partial derivative matrices of  
 826 the  $f_s$ 's with respect to interval spanned by  $s$ , as sub-matrices of  
 827  $H$ . If one can find a subset  $A \subseteq [d]$  such that each  $s$  spans an  
 828 interval  $I(s)$  that is  $\delta$ -close to  $A$  for some small  $\delta$ , then one obtains  
 829 a lower bound for polynomials  $f$  with high rank with respect to  $A$ .

830 This first method leads to Theorem 4.1 and it is already strong  
 831 enough to prove several lower bounds. We believe that in many oc-  
 832 currences in the literature, when obtaining lower bounds involving  
 833 a circuit decomposition and a partial derivative matrix with respect  
 834 to a given partition of the set of positions  $[d]$ , this is somehow the  
 835 underlying method.

836 However, this method poorly makes use of the structure of  $H$ ,  
 837 since it may happen that some of the chosen sub-blocks are face to  
 838 face with one another. A short illustration of this phenomenon is  
 839 the following. Let

$$M = \left( \begin{array}{cc|cc} A_{1,1} & A_{1,2} & & \\ A_{2,1} & A_{2,2} & C_1 & \\ \hline & C_2 & B_{1,1} & B_{1,2} \\ & & B_{2,1} & B_{2,2} \end{array} \right)$$

840 be a block matrix, for which one wants to obtain a lower bound  
 841 on the rank, knowing a lower bound on rank  $\left( \sum_{i,j} A_{i,j} + B_{i,j} \right)$ , and  
 842 with no assumption on the  $C_i$ 's.

843 The previous method would go as follows:

$$\begin{aligned} \text{rank}(M) &\geq \max \left[ \max_{i,j} \text{rank}(A_{i,j}), \max_{i,j} \text{rank}(B_{i,j}) \right] \\ &\geq \frac{1}{8} \sum_{i,j} \text{rank}(A_{i,j}) + \text{rank}(B_{i,j}) \\ &\geq \frac{1}{8} \text{rank} \left( \sum_{i,j} A_{i,j} + B_{i,j} \right). \end{aligned}$$

844 Note that we have lost a factor of 8, which is the number of small  
 845 blocks that we wish to sum.

846 A more efficient method would consist in first summing rows  
 847 and columns of  $M$  in order to put together the  $A$ 's and the  $B$ 's.

848 This would go as follows, for some matrices  $C'_1$  and  $C'_2$ ,

$$\begin{aligned} \text{rank}(M) &\geq \text{rank} \left( \begin{bmatrix} \sum_{i,j} A_{i,j} & C'_1 \\ C'_2 & \sum_{i,j} B_{i,j} \end{bmatrix} \right) \\ &\geq \max \left[ \text{rank} \left( \sum_{i,j} A_{i,j} \right), \text{rank} \left( \sum_{i,j} B_{i,j} \right) \right] \\ &\geq \frac{1}{2} \text{rank} \left( \sum_{i,j} A_{i,j} + B_{i,j} \right). \end{aligned}$$

849 By doing so, we have decreased the factor 8 to 2, which is the  
850 number of larger blocks.

851 Back to the Hankel matrix  $H$ , this corresponds to putting  
852 together the polynomials  $f_s$  for which we have chosen the same  
853 spanned interval (this corresponds to  $d^2$  larger blocks) instead of  
854 considering them separately (which corresponds to  $|\text{PT}(\mathcal{C})|$  smaller  
855 blocks). We now formalize this idea, using a total order to model  
856 the choice of intervals for convenience.

857 LEMMA 4.4. *Let  $\tilde{f} : \text{Tree}(X) \rightarrow K$  be a non-associative non-*  
858 *commutative polynomial and let  $\leq_{\text{int}}$  be a total order on inter-*  
859 *vals of  $[d]$ . For any shape  $s$ , let  $I(s)$  be the smallest (with respect*  
860 *to  $\leq_{\text{int}}$ ) interval spanned by  $s$ . For any interval  $I$ , define a non-*  
861 *commutative associative polynomial by*

$$\begin{aligned} f_I : X^* &\rightarrow K \\ u &\mapsto \sum_{\substack{t \in \text{Tree}(X) \\ \text{label}(t)=u \\ I(\text{shape}(t))=I}} \tilde{f}(t). \end{aligned}$$

862 Then, one has  $\max_I \text{rank}(M_I(f_I)) \leq \text{rank}(H_{\tilde{f}})$ .

863 We illustrate the definition of  $f_I$  through a small example. Let  
864  $t = ((xy)z)$ , and assume  $[1, 2]$  is the smallest interval spanned by  
865  $t$ , that is,  $[1, 2] \leq_{\text{int}} \{1\}, \{2\}, \{3\}, [1, 3]$ . Then  $\tilde{f}(t)$  will contribute  
866 to  $f_{[1,2]}(xyz)$  as  $\text{label}(t) = xyz$  and  $I(\text{shape}(t)) = [1, 2]$ .

867 PROOF. Our aim is to obtain  $M_I(f_I)$  from  $H_{\tilde{f}}$ , by first taking a  
868 sub-matrix, then adequately summing its rows and columns. The  
869 proof is summarized in Figure 4.2.

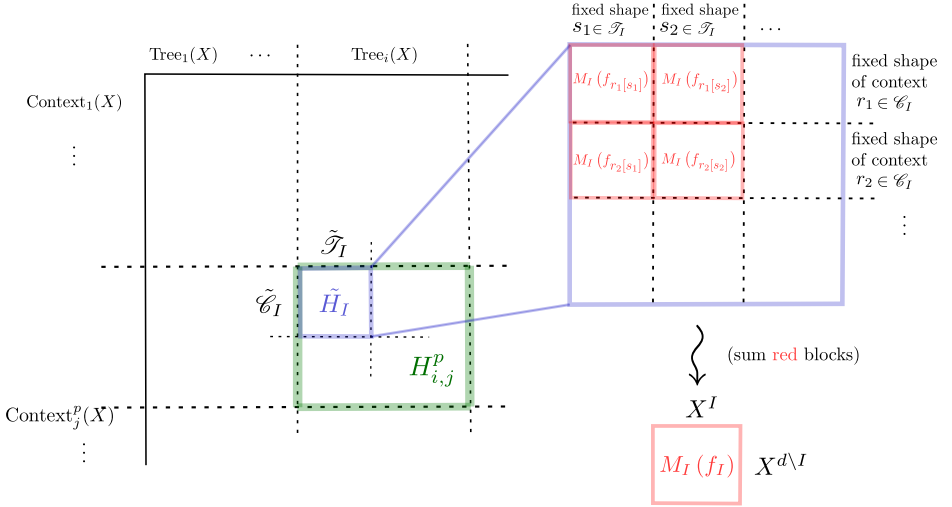


Figure 4.2: Decomposition of the Hankel matrix used in the proof of Lemma 4.4. Here,  $I = [p, p + i - 1]$ .

870 Let  $I = [p, p + i - 1]$  be some fixed interval and  $j = d - i + 1$ .

871 Let  $r \in \text{Context}_j^p$  be a shape of a context of size  $j$  and where  $\square$   
 872 is on the  $p$ -th leaf, let  $v$  be a node in  $r$  and let  $[a, b]$  be the interval  
 873 spanned by  $v$  in  $r$ . We define the interval  $I'_v$  by

$$I'_v = \begin{cases} [a, b] & \text{if } b < p \\ [a, b + i - 1] & \text{if } a \leq p \leq b \\ [a + i - 1, b + i - 1] & \text{if } a > p, \end{cases}$$

874 The interval  $I'_v$  is to be seen as the interval of positions of the leaves  
 875 that are descendants of  $v$  in some  $r[s']$  where  $s'$  is any element of  
 876  $\text{Tree}_i$ . In particular, if  $v$  is the leaf labelled by  $\square$  in  $r$ , then  $I'_v = I$ .

877 Likewise, for a node  $v$  of a (sub)shape  $s' \in \text{Tree}_i$ , we define  $I'_v$   
 878 by  $I'_v = [a + p - 1, b + p - 1]$ , where  $[a, b]$  is the interval spanned  
 879 by  $v$  in  $s'$ . Note that if  $v$  is the root of  $s'$  then  $I_v = I$ .

880 We may now define (we use order  $\leq_{int}$  on intervals)

$$\mathcal{C}_I = \{r \in \text{Context}_j^p \mid I = \min_{v \text{ node in } r} I'_v\},$$

881 and

$$\mathcal{T}_I = \{s' \in \text{Tree}_i \mid I = \min_{v \text{ node in } s'} I'_v\}.$$

882 We extend these subsets to labelled trees and context in a  
 883 straightforward fashion by defining  $\tilde{\mathcal{C}}_I = \{c \in \text{Context}_j^p(X) \mid$   
 884  $\text{shape}(c) \in \mathcal{C}_I\}$  and  $\tilde{\mathcal{T}}_I = \{t \in \text{Tree}_i(X) \mid \text{shape}(t) \in \mathcal{T}_I\}$ .

885 Remark that for any  $t \in \text{Tree}(X)$  and  $u \in X^*$ , one has  $\text{label}(t) =$   
 886  $u$  and  $I = I(\text{shape}(t))$  if and only if  $t = r[s]$  for some  $(s, r) \in$   
 887  $\tilde{\mathcal{T}}_I \times \tilde{\mathcal{C}}_I$  such that  $u = \text{label}(s) \otimes_I \text{label}(c)$ .

888 We now consider the submatrix  $\tilde{H}_I$  of  $H_{i,j}^p$  where the rows are  
 889 restricted to  $\tilde{\mathcal{C}}_I$  and the columns to  $\tilde{\mathcal{T}}_I$ . In this matrix, we now sum  
 890 the rows which are indexed by contexts with the same label, and the  
 891 columns which are indexed by trees with the same label, to obtain  
 892 matrix  $H_I$ . Clearly,  $\text{rank}(H_I) \leq \text{rank}(H_{\tilde{f}})$ . We finally prove that  
 893  $H_I = M_I(f_I)$ . Indeed, let  $g \in X^I \simeq X^i$  and  $h \in X^{d \setminus A} \simeq X^j$ . Then

$$M_I(f_I)(g, h) = \sum_{\substack{t \in \text{Tree}(X) \\ \text{label}(t) = g \otimes_I h \\ I(\text{shape}(t)) = I}} \tilde{f}(t) = \sum_{\substack{s \in \tilde{\mathcal{T}}_I \\ c \in \tilde{\mathcal{C}}_I \\ \text{label}(s) = g \\ \text{label}(c) = h}} \tilde{f}(c[s]) = H_I(g, h),$$

894 which concludes the proof of Lemma 4.4. □

895 With Lemma 4.4 in hands, we are ready to prove Theorem 4.2.  
 896 Let  $\tilde{f} : \text{Tree}(X) \rightarrow K$  be the non-associative polynomial computed  
 897 by  $\mathcal{C}$  when seen as a non-associative circuit. Let  $\leq_{int}$  be a total  
 898 order on intervals of  $d$  such that  $I \mapsto \text{dist}(I, A)$  is non-decreasing.  
 899 In other words,  $I_1 <_{int} I_2$  if and only if  $d(I_1, A) < d(I_2, A)$ . Let  
 900  $f_I : X^d \rightarrow K$  be given by

$$f_I(u) = \sum_{\substack{t \in \text{Tree}(X) \\ \text{label}(t) = u \\ I(\text{shape}(t)) = I}} \tilde{f}(t).$$

901 Then any interval  $I$  such that  $d(I, A) > \delta$  is such that for every  
 902 parse tree  $s \in \text{PT}(\mathcal{C})$ , one has  $I \neq I(s)$ , so  $f_I = 0$ . Hence, we  
 903 obtain

$$\begin{aligned}
 \text{rank}(M_A(f)) &= \text{rank} \left( M_A \left( \sum_{I \text{ interval of } [d]} f_I \right) \right) \\
 &= \text{rank} \left( M_A \left( \sum_{\substack{I \text{ interval of } [d] \\ \text{dist}(A,I) \leq \delta}} f_I \right) \right) \\
 &\leq \sum_{\substack{I \text{ interval of } [d] \\ \text{dist}(A,I) \leq \delta}} \text{rank}(M_A(f_I)) && \text{by rank subadditivity} \\
 &\leq \sum_{\substack{I \text{ interval of } [d] \\ \text{dist}(A,I) \leq \delta}} n^\delta \text{rank}(M_I(f_I)) && \text{by Lemma 3.8} \\
 &\leq d^2 n^\delta \text{rank}(H_{\bar{f}}) && \text{by Lemma 4.4}
 \end{aligned}$$

904 which yields the announced lower bound.

905 **4.3. General Lower Bounds in the Commutative Setting.**

906 We explain how to extend the notions of parse trees and the generic  
 907 lower bound theorems to the associative commutative setting.

908 Let  $X = X_1 \sqcup X_2 \sqcup \dots \sqcup X_d$  be a partition of the set  $X$  of  
 909 variables. A monomial is *set-multilinear* with respect to the parti-  
 910 tion if it is the product of exactly one variable from each set  $X_i$ ,  
 911 and a polynomial is set-multilinear if all its monomials are.

912 **EXAMPLE 4.5.** The permanent and the determinant of degree  $d$   
 913 are set-multilinear with respect to the partition  $X = X_1 \sqcup X_2 \sqcup \dots \sqcup$   
 914  $X_d$  where  $X_i = \{x_{i,j}, j \in [d]\}$ . The iterated matrix multiplication  
 915 polynomial is another example of an important and well-studied  
 916 set-multilinear polynomial.  $\diamond$

917 Partial derivative matrices also make sense in the realm of set-  
 918 multilinear polynomials.

919 **DEFINITION 4.6.** Let  $X = X_1 \sqcup X_2 \sqcup \dots \sqcup X_d$ ,  $f$  be a set-multilinear  
 920 polynomial of degree  $d$ , and  $A \subseteq [d]$  be a set of indices. The *partial*  
 921 *derivative matrix*  $M_A(f)$  of  $f$  with respect to  $A$  is defined as

922 follows: the rows are indexed by set-multilinear monomials  $g$  with  
 923 respect to the partition  $\bigsqcup_{i \notin A} X_i$  and the columns are indexed by  
 924 set-multilinear monomials  $h$  with respect to the partition  $\bigsqcup_{i \in A} X_i$ .  
 925 The value of  $M_A(f)(g, h)$  is the coefficient of the monomial  $g \cdot h$   
 926 in  $f$ .

927 The notion of shape was defined by (Arvind & Raja 2016), and  
 928 it slightly differs from the non-commutative setting because we  
 929 need to keep track of the indices of the variable sets given by the  
 930 partition from which the variables belong. More precisely, given a  
 931 partition of  $X = X_1 \sqcup X_2 \sqcup \dots \sqcup X_d$ , we associate to any monomial  
 932  $t \in \text{Tree}(X)$  of degree  $d$  its **shape**  $\text{shape}(t) \in \text{Tree}_d([d])$  defined as  
 933 the tree obtained from  $t$  by replacing each label by its index in the  
 934 partition. In particular if  $t$  is set-multilinear, then each element  
 935 in  $[d]$  appears exactly once as an index in  $\text{shape}(t)$ . Hence we let  
 936  $\mathcal{T}_d \subseteq \text{Tree}_d([d])$  denote the set of trees such that all elements of  $[d]$   
 937 appear exactly once as a label of a leaf.

Let  $\mathcal{C}$  be a commutative circuit. We let  $\tilde{f}$  denote the commu-  
 tative non-associative polynomial computed by  $\mathcal{C}$  when it is seen  
 as non-associative. A **parse tree** of  $\mathcal{C}$  is any shape  $s \in \mathcal{T}_d$  for  
 which there exists a monomial  $t \in \text{Tree}(X)$  whose coefficient in  $\tilde{f}$   
 is non-zero and such that  $s = \text{shape}(t)$ . Formally, we let

$$\text{PT}(\mathcal{C}) = \left\{ \text{shape}(t) \mid \tilde{f}(t) \text{ non-zero} \right\} \cap \mathcal{T}_d.$$

938 REMARK 4.7. Note that it may be the case that a circuit  $\mathcal{C}$  com-  
 939 puting a set-multilinear polynomial  $f$  computes a non-associative  
 940  $\tilde{f}$  such that  $\tilde{f}(t) \neq 0$  for some non set-multilinear monomials  $t$ ,  
 941 provided their sums collapse to 0 in the associative setting. We  
 942 do not count such shapes as parse trees (this explains the intersec-  
 943 tion with  $\mathcal{T}_d$  in the above definition), which leads to more general  
 944 classes of circuits against which we shall obtain lower bounds.

945 Given a shape  $s \in \mathcal{T}_d$  and a node  $v$  of  $s$ , we let  $s_v$  denote the  
 946 subtree rooted at  $v$  and  $A_v \subseteq [d]$  denote the set of labels appearing  
 947 on the leaves of  $s_v$ . We say that  $A_v$  is **spanned** by  $s$ .

948 Following the same roadmap as in the non-commutative setting  
 949 we obtain the following counterpart of Theorem 4.1. We assume



950 that the set of variables is partitioned into  $d$  parts of equal size  $n$   
 951 (this is a natural setting for polynomials such as the determinant,  
 952 the permanent or the iterated matrix multiplication). In particu-  
 953 lar, it means that the polynomials we consider are of degree  $d$  and  
 954 over  $nd$  variables.

955 THEOREM 4.8. *Let  $f$  be a set-multilinear polynomial computed*  
 956 *by a circuit  $\mathcal{C}$ . Let  $A \subseteq [d]$  and  $\delta \in \mathbb{N}$  such that all parse trees of*  
 957  *$\mathcal{C}$  span a subset at distance at most  $\delta$  from  $A$ . Then  $\mathcal{C}$  has size at*  
 958 *least  $\text{rank}(M_A(f)) n^{-\delta} |\text{PT}(\mathcal{C})|^{-1}$ .*

959 PROOF. As this proof is an adaptation of that of Theorem 4.1,  
 960 we concentrate on highlighting the necessary changes.

Let  $X = X_1 \sqcup X_2 \sqcup \dots \sqcup X_d$  denote the underlying partition. Previously, we grouped together (sub-)trees and (sub-)contexts which correspond to a given interval of positions. In the commutative setting, we instead group together the (sub-)trees and (sub-)contexts which correspond to a given *subset* of positions, where a position is now being given by its index in the partition. Formally, for  $A \subseteq [d]$ , we let

$$\text{Tree}_A(X) = \{t \in \text{Tree}(X) \mid \begin{array}{l} \text{the set of indices of variables} \\ \text{labeling } t \text{ is } A \}, \end{array}$$

and likewise,

$$\text{Context}_A(X) = \{c \in \text{Context}(X) \mid \begin{array}{l} \text{the set of indices of variables} \\ \text{(different from } \square) \text{ labeling } c \text{ is } A \}, \end{array}$$

and finally

$$H_A = H_{|\text{Tree}_A(X) \times \text{Context}_A(X)}.$$

961 Now, grouping together the columns of  $H_A$  which correspond to  
 962 trees which have a given fixed shape  $s'$  (recall that a commutative  
 963 shape contains the index in the partition of each leaf), and the  
 964 rows which correspond to contexts which have a given fixed shape  
 965 of context  $r$  yields the partial derivative matrix  $M_A(f_{r[s']})$ , where

966 the (commutative, associative) polynomial  $f_s$  is defined, for any  
 967 commutative shape  $s$ , by

$$f_s(u) = \tilde{f}(s \text{ labelled by } u),$$

968 where the labeling respects the partition of  $X$ .

969 Hence,  $\text{rank}(H) \geq \text{rank}(M_A(f_s))$  whenever  $A$  is spanned by  
 970  $s$ . The remainder of the proof exactly follows that of Theorem 4.1  
 971 and therefore we do not repeat it here.  $\square$

972 A notable difference with the non-commutative setting is that  
 973 now parse trees no longer span intervals of  $[d]$  but subsets of  $[d]$ . As  
 974 a consequence, if we follow the same technique as the one used to  
 975 prove Theorem 4.2, we now groups together blocks corresponding  
 976 to the same *subset* of  $[d]$  and therefore the multiplicative factor is  
 977 now  $2^{-d}$  as there are  $2^d$  such subsets. This yields the following  
 978 counterpart for Theorem 4.2.

979 **THEOREM 4.9.** *Let  $f$  be a set-multilinear polynomial computed*  
 980 *by a circuit  $\mathcal{C}$ . Let  $A \subseteq [d]$  and  $\delta \in \mathbb{N}$  such that all parse trees of*  
 981  *$\mathcal{C}$  span a subset at distance at most  $\delta$  from  $A$ . Then  $\mathcal{C}$  has size at*  
 982 *least  $\text{rank}(M_A(f)) n^{-\delta} 2^{-d}$ .*

983 **PROOF.** Again, we extend the ideas for the non-commutative  
 984 setting to the commutative setting, and we reuse the notations of  
 985 the proof of Theorem 4.2. As for proving Theorem 4.2, we mainly  
 986 rely on a Lemma.

987 **LEMMA 4.10.** *Let  $\tilde{f} : \text{Tree}(X) \rightarrow K$  be a non-associative com-*  
 988 *mutative polynomial and let  $\leq_{int}$  be a total order on subsets of*  
 989  *$[d]$ . For any commutative shape  $s$ , let  $A(s)$  be the smallest (with*  
 990 *respect to  $\leq_{int}$ ) subset spanned by  $s$ . For any subset  $A$ , define a*  
 991 *commutative associative polynomial by*

$$f_A(u) = \sum_{\substack{t \in \text{Tree}(X) \\ \text{label}(t)=u \\ A(\text{shape}(t))=A}} \tilde{f}(t).$$

992 Then, one has  $\max_A \text{rank}(M_A(f_A)) \leq \text{rank}(H_{\tilde{f}})$ .

The proof of Lemma 4.10 is very similar, yet a bit more pleasant than that of Lemma 4.4, since we no longer need to shift any interval. Formally, for  $A \subseteq [d]$  we define

$$\mathcal{T}_A = \{t \in \text{Tree}_A(X) \mid A \text{ is the smallest interval spanned by } \text{shape}(t)\},$$

and likewise,

$$\mathcal{C}_A = \{c \in \text{Context}_A(X) \mid A \text{ is the smallest interval spanned by } \text{shape}(c)\}.$$

993 Now, the lemma follows from the fact that  $M_A(f_A)$  is obtained by  
994 summing rows from  $\mathcal{T}_A$  and columns from  $\mathcal{C}_A$  in  $H$ .

995 The remainder of the proof is a very straightforward adaptation  
996 of the end of the proof of Theorem 4.2 from the non-commutative  
997 to the commutative setting.  $\square$

998 **REMARK 4.11.** *While in the non-commutative setting, Theorem 4.2*  
999 *strengthens Theorem 4.1 (when  $d^2$  is small), this is no longer the*  
1000 *case in the commutative setting. Indeed, the maximal number of*  
1001 *commutative parse trees being roughly  $d!$  (the exact asymptotic is*  
1002  $\frac{\sqrt{2-\sqrt{2}}d^{d-1}}{e^d(\sqrt{2-1})^{d+1}}$ *, see Sloane (2011)), Theorem 4.8 and Theorem 4.9 are*  
1003 *incomparable.*

## 1004 5. Applications

1005 In this section we instantiate our generic lower bound theorems on  
1006 concrete classes of circuits. We first show how the weaker version  
1007 (Theorem 4.1) yields the best lower bounds to date for skew and  
1008 small non-skew depth circuits. Extending these ideas we obtain  
1009 exponential lower bounds for  $(\frac{1}{2} - \varepsilon)$ -unbalanced circuits, an ex-  
1010 tension of skew circuits which are just slightly unbalanced. We also  
1011 adapt the proof to  $\varepsilon$ -balanced circuits, which are slightly balanced.  
1012 We then move on to our main results, which concern circuits with  
1013 many parse trees, with lower bounds for both non-commutative  
1014 and commutative settings.

1015 Prior to that, we present a family of polynomials for which our  
 1016 lower bounds hold, and we state Lemma 5.1 which is used several  
 1017 times in our proofs.

1018 **High-Ranked Polynomials.** The lower bounds we state below  
 1019 hold for any polynomial whose partial derivative matrices with  
 1020 respect to either a fixed subset  $A$  or all subsets have full rank.  
 1021 Such polynomials exist for all fields in both the commutative and  
 1022 non-commutative settings, and can be explicitly constructed. For  
 1023 instance the so-called Nisan-Wigderson polynomial (Kayal *et al.*  
 1024 2014b) — inspired by the notion of designs by Nisan and Wigder-  
 1025 son (Nisan & Wigderson 1994) — has this property. In the com-  
 1026 mutative, set-multilinear setting, it is given by

$$NW_{n,d} = \sum_{\substack{h \in \mathbb{F}_n[z] \\ \deg(h) \leq d/2}} \prod_{i=1}^d x_{i,h(i)},$$

1027 where  $\mathbb{F}_n[z]$  denotes univariate polynomials with coefficients in the  
 1028 finite field of prime order  $n$ . In the non-commutative setting, we  
 1029 remove index  $i$ , and insist that the product  $\prod_{i=1}^d x_{h(i)}$  is done along  
 1030 increasing values of  $i$ . The fact that there exists a unique polyno-  
 1031 mial  $h \in \mathbb{F}_n[z]$  of degree at most  $d/2$  which takes  $d/2$  given values  
 1032 at  $d/2$  given positions exactly implies that the partial derivative  
 1033 matrix of  $NW_{n,d}$  with respect to any  $A \subseteq [d]$  of size  $d/2$  is a per-  
 1034 mutation matrix. This is then easily extended to any  $A \subseteq [d]$ .

1035  **$A$ -balanced subsets.** The following combinatorial Lemma is widely  
 1036 used to derive our lower bounds. Intuitively, a subset  $B \subseteq [d]$  is  
 1037 far from a subset  $A \subseteq [d]$  of size  $d/2$  whenever it is  $A$ -balanced,  
 1038 meaning that  $A \cap B$  and  $A^c \cap B$  have roughly the same size.

LEMMA 5.1. *Let  $A, B \subseteq [d]$  be such that  $|A| = d/2$ . Then*

$$d(A, B) = d/2 - \left| |A \cap B| - |A^c \cap B| \right|.$$

1039 PROOF. Let us first assume that  $|B \cap A| \geq |B|/2$ . This implies

1040 that  $|\Delta(A, B)| \leq |\Delta(A^c, B)|$ , so

$$\begin{aligned}
 \text{dist}(A, B) &= |\Delta(A, B)| \\
 &= |A \cup B| - |A \cap B| \\
 &= (|A| + |A^c \cap B|) - |A \cap B| \\
 &= d/2 - (|A \cap B| - |A^c \cap B|) \\
 &= d/2 - \left| |A \cap B| - |A^c \cap B| \right|,
 \end{aligned}$$

1041 where the last line also follows from the assumption that  $|B \cap A| \geq$   
 1042  $|B|/2$ . Now if  $|B \cap A| < |B|/2$ , it suffices to replace  $A$  with  $A^c$  in  
 1043 the previous proof to obtain the announced result.  $\square$

## 1044 5.1. Applications in the non-commutative setting.

1045 **5.1.1. Skew, Slightly Unbalanced, Slightly Balanced and**  
 1046 **Small Non-Skew Depth Circuits.** We show how using Theo-  
 1047 rem 4.1 yields exponential lower bounds for four classes of circuits  
 1048 in the non-commutative setting. We adapt the ideas of (Limaye  
 1049 *et al.* 2016) into our newly introduced vocabulary and easily obtain  
 1050 the same exponential lower bounds for skew circuits. Straightfor-  
 1051 ward generalizations lead to previously unknown exponential lower  
 1052 bounds on slightly unbalanced and slightly balanced circuits. Fi-  
 1053 nally, we also adapt (and shorten) their proof of a lower bound on  
 1054 small non-skew depth circuits. In each of these four cases the use  
 1055 of our weaker theorem, namely Theorem 4.1 suffices.

1056 **Skew Circuits** A circuit  $\mathcal{C}$  is *skew* if all its parse trees are skew,  
 1057 meaning that each node has at least one of its children which is  
 1058 a leaf. We let  $I_{mid} = (d/4, 3d/4]$ , which has size  $d/2$ . As a direct  
 1059 application of Theorem 4.1, we obtain the following result.

1060 **THEOREM 5.2.** *Let  $f$  be a homogeneous non-commutative poly-*  
 1061 *nomial of degree  $d$  and on  $n$  variables such that  $M_{I_{mid}}(f)$  has full*  
 1062 *rank  $n^{d/2}$ . Then any skew circuit computing  $f$  has size at least*  
 1063  *$2^{-d}n^{d/4}$ .*

1064 **PROOF.** The proof relies on the following two easy observations.

1065 FACT 5.3. Any skew shape spans intervals of each possible size,  
 1066 and in particular, an interval of size  $3d/4$ .

1067 PROOF. Let  $s \in \text{Tree}_d$  be a skew shape,  $v_1$  be its root, and for  
 1068 all  $i = 1 \dots d - 2$ ,  $v_{i+1}$  be the child of  $v_i$  which is not a leaf. Then  
 1069 any of the two children of  $v_{d-2}$  is a leaf, so it spans an interval of  
 1070 size 1. Now for each  $i$ ,  $v_i$  spans an interval that includes  $I_{v_{i+1}}$  and  
 1071 adds 1 to its size, so we easily conclude by induction.  $\square$

1072 FACT 5.4. Any interval of size  $3d/4$  is at distance at most (in fact,  
 1073 equal to)  $d/4$  from  $I_{mid}$ .

1074 PROOF. Indeed, let  $I \subseteq [d]$  be an interval of size  $3d/4$ . Then  
 1075  $I_{mid} \subseteq I$  (see Figure 5.1). Hence by Lemma 5.1,

$$\begin{aligned} d(I, I_{mid}) &= d/2 - \left| |I \cap I_{mid}| - |I \cap I_{mid}^c| \right| \\ &= d/2 - |d/2 - (|I| - d/2)| = d/4. \end{aligned}$$

1076  $\square$

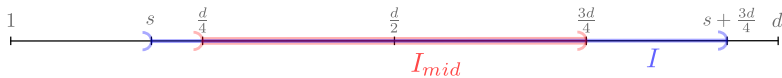


Figure 5.1: Any interval  $I$  of size  $\frac{3d}{4}$  is at distance  $\frac{d}{4}$  from  $I_{mid}$ .

1077 A skew circuit has only skew parse trees, which all span an  
 1078 interval of size  $3d/4$ . Such an interval is at distance  $d/4$  from  $I_{mid}$ ,  
 1079 so the announced lower bound follows directly from Theorem 4.1,  
 1080 together with the fact that there are  $2^d$  skew trees.  $\square$

1081 REMARK 5.5. Note that the factor  $2^{-d}$  is easily replaced by  $d^{-2}$   
 1082 by applying Theorem 4.2 instead, but we find it remarkable that  
 1083 simply using a decomposition of  $H$  into blocks is enough to obtain  
 1084 such an exponential lower bound.

1085 **Slightly Unbalanced Circuits** A circuit  $\mathcal{C}$  computing a ho-  
 1086 mogeneous non-commutative polynomial of degree  $d$  is said to be  
 1087  **$\alpha$ -unbalanced** if every multiplication gate has at least one of its  
 1088 children which computes a polynomial of degree at most  $\alpha d$ .

1089 THEOREM 5.6. *Let  $f$  be a homogeneous non-commutative poly-*  
 1090 *nomial of degree  $d$  and on  $n$  variables such that  $M_{I_{mid}}(f)$  has full*  
 1091 *rank  $n^{d/2}$ . Then any  $(\frac{1}{2} - \varepsilon)$ -unbalanced circuit computing  $f$  has*  
 1092 *size at least  $4^{-d}n^{\varepsilon d}$ .*

1093 This result improves over a previously known exponential lower  
 1094 bound on  $(\frac{1}{5})$ -unbalanced circuits (Limaye *et al.* 2016).

1095 PROOF. This is an adaptation of the proof of Theorem 5.2 about  
 1096 skew circuits. We now rely on these two observations, which re-  
 1097 spectively generalize Fact 5.3 and Fact 5.4:

1098 FACT 5.7. *Any  $(\frac{1}{2} - \varepsilon)$ -unbalanced shape spans an interval of size*  
 1099 *between  $3d/4 - (\frac{1}{2} - \varepsilon)d/2$  and  $3d/4 + (\frac{1}{2} - \varepsilon)d/2$ , that is, between*  
 1100  *$d/2 + d\varepsilon/2$  and  $d - d\varepsilon/2$ .*

1101 PROOF. Let  $\alpha$  denote  $(\frac{1}{2} - \varepsilon) < 1/2$  and let  $s \in \text{Tree}_d$  be an  
 1102  $\alpha$ -unbalanced shape of size  $d$ . We let  $v_1$  be its root, and  $v_2$  be  
 1103 the child of  $v_1$  which spans the largest interval, which has size  
 1104  $I_{v_2} \geq (1 - \alpha)d \geq \alpha d$ . If both children of  $I_{v_2}$  span intervals of size  
 1105  $\leq \alpha d$ , we set  $r = 2$ , and otherwise iterate for  $i = 3 \dots r$  until both  
 1106 children of  $v_r$  span intervals of size  $\leq \alpha d$ . Now, if we choose  $v_{r+1}$   
 1107 to be a child of  $v_r$ , the cardinalities of the growing sequence of  
 1108 intervals  $I_{v_{r+1}} \subseteq I_{v_r} \subseteq \dots \subseteq I_{v_1} = [d]$  range from  $\leq \alpha d$  to  $d$  with  
 1109 differences bounded by  $\alpha d$ , so one of the interval has a size lying  
 1110 in  $[3d/4 - \alpha/2, 3d/4 + \alpha/2]$ .  $\square$

1111 FACT 5.8. *Any interval  $I$  of size  $d/2 + \varepsilon d/2 \leq |I| \leq d - \varepsilon d/2$  is at*  
 1112 *distance  $\leq d/2 - \varepsilon d/2$  from  $I_{mid}$ .*

1113 PROOF. We make a case distinction and first assume that  $I_{mid} \subseteq$

1114  $I$ . Then, by Lemma 5.1, we have that

$$\begin{aligned} \text{dist}(I, I_{mid}) &= d/2 - \left| |I \cap I_{mid}| - |I \cap I_{mid}^c| \right| \\ &= d/2 - |d/2 - (|I| - d/2)| \\ &= d - |I| < d/2 - \varepsilon d/2. \end{aligned}$$

1115 Assume now that  $I_{mid} \not\subseteq I$ . Then, either  $3d/4$  or  $d/4 + 1$  does not  
 1116 belong to  $I$ . Both cases being symmetrical, we assume without  
 1117 loss of generality that  $3d/4 \notin I$ . We let  $\ell = |I \cap I_{mid}|$ . The current  
 1118 situation is depicted in Figure 5.2.

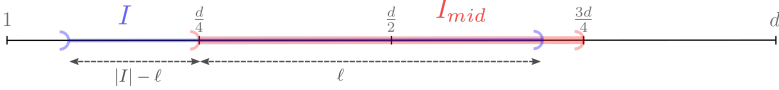


Figure 5.2: Illustrating  $I$  and  $I_{mid}$  when  $3d/4 \notin I$ .

1119 It follows that  $|I \cap [3d/4, d]| = 0$ , and  $|I \cap [1, d/4]| = (|I| - \ell) \leq$   
 1120  $d/4$ . Multiplying this last inequality by two and summing with  
 1121  $-|I| \leq -d/2$  yields  $|I| - 2\ell \leq 0$ , so we obtain

$$\begin{aligned} \text{dist}(I, I_{mid}) &= d/2 - \left| |I \cap I_{mid}| - |I \cap I_{mid}^c| \right| \\ &= d/2 - \left| \ell - (|I \cap [1, d/4]| + |I \cap [3d/4, d]|) \right| \\ &= d/2 - \left| \ell - (|I| - \ell) \right| \\ &= d/2 - (2\ell - |I|). \end{aligned}$$

1122 Now since  $|I| - \ell \leq d/4$ ,  $-2\ell \leq -2|I| + d/2$ , which leads to

$$\begin{aligned} \text{dist}(I, I_{mid}) &= d/2 - 2\ell + |I| \\ &\leq d - |I| \leq d/2 - \varepsilon d/2, \end{aligned}$$

1123 since  $|I| \geq d/2 + \varepsilon d/2$ . □

1124 We conclude the proof by applying Theorem 4.1, just as we did  
 1125 for skew circuits. □



1126 **Slightly Balanced Circuits** A circuit  $\mathcal{C}$  computing a homo-  
 1127 geneous non-commutative polynomial of degree  $d$  is said to be  $\alpha$ -  
 1128 **balanced** if every multiplication gate which computes a polynomial  
 1129 of degree  $k$  has both of its children which compute polynomials of  
 1130 degree at least  $\alpha k$ .

1131 **THEOREM 5.9.** *Let  $f$  be a homogeneous non-commutative poly-*  
 1132 *nomial of degree  $d$  and on  $n$  variables such that  $M_{[1,d/2]}(f)$  has full*  
 1133 *rank  $n^{d/2}$ . Then any  $\varepsilon$ -balanced circuit computing  $f$  has size at*  
 1134 *least  $4^{-d}n^{\varepsilon d}$ .*

1135 **PROOF.** Let  $s$  be an  $\varepsilon$ -balanced shape, and  $r$  be the root of  $s$ .  
 1136 Let  $I = [1, b]$  be the interval spanned by the left child of  $r$ . Since  $s$   
 1137 is  $\varepsilon$ -balanced,  $\varepsilon d \leq |I| = b \leq (1 - \varepsilon)d$ . Hence,  $I$  is at a distance of  
 1138 at most  $d/2 - \varepsilon d$  from  $[1, d/2]$ , which allows us to conclude using  
 1139 Theorem 4.1.  $\square$

1140 Note that it suffices to simply restrict the *last* multiplication in  
 1141 the circuit to be  $\varepsilon$ -balanced for the proof to carry on.

1142 **Small Non-Skew Depth Circuits** A circuit  $\mathcal{C}$  has **non-skew**  
 1143 **depth  $k$**  if all its parse trees are such that each path from the root  
 1144 to a leaf goes through at most  $k$  non-skew nodes, i.e., nodes for  
 1145 which the two children are inner nodes. We obtain an alternative  
 1146 proof of the exponential lower bound of (Limaye *et al.* 2016) on  
 1147 non-skew depth  $k$  circuits as an application of Theorem 4.1. In  
 1148 the rest of this section we assume that  $k \geq 30$ ,  $p \geq 30$  is some  
 1149 multiple of 3 and  $d = 12kp$ . We will make extended use of the  
 1150 subset  $A \subseteq [d]$  introduced in (Limaye *et al.* 2016),

$$A = [1, 3kp] \cup \bigcup_{i=1}^{3k} [3(k+i)p + 2p, 3(k+i+1)p] \subseteq [d],$$

1151 of size  $6kp = d/2$  which is better understood in Figure 5.3.

1152 **THEOREM 5.10.** *Let  $f$  be a homogeneous non-commutative poly-*  
 1153 *nomial of degree  $d = 12kp$  and on  $n$  variables such that  $M_A(f)$  has*  
 1154 *full rank  $n^{d/2}$ . Then any circuit of non-skew depth  $k$  computing  $f$*   
 1155 *has size at least  $4^{-d}n^{p/3} = 4^{-d}n^{d/36k}$ .*

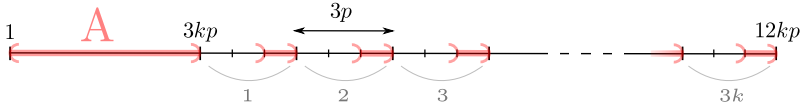


Figure 5.3: Subset  $A \subseteq [d]$ .

1156 PROOF. We shall prove that any parse tree  $s \in \text{Tree}_d$  with non-  
 1157 skew depth  $k$  spans an interval  $I(s)$  at distance  $\leq d/2 - p/3$  from  
 1158  $A$ . Then the result follows by applying Theorem 4.1.

1159 Assume towards contradiction that a non-skew depth  $k$  shape  
 1160  $s \in \text{Tree}_d$  spans only interval at distance  $> d/2 - p/3$  from  $A$ . We  
 1161 consider (see Figure 5.4) the path  $v_1 \cdots v_r$  in  $s$  from its root to the  
 1162 leaf with position  $3kp$ , and write  $u_i$  for  $i \in r - 1$ , to refer to the  
 1163 child of  $v_i$  which is not  $v_{i+1}$ . Since  $s$  has non-skew depth  $k$ , at least  
 1164  $r - k$  nodes among  $v_1, \dots, v_{r-1}$  are leaves.

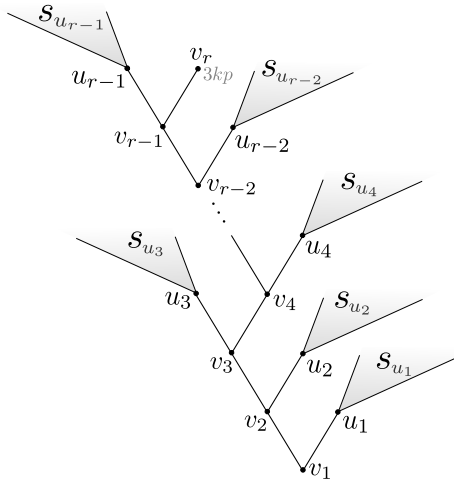


Figure 5.4: The path from the root  $v_1$  to  $v_r$ , the leaf with position  $3kp$ .

1165 We now state and prove some facts which then lead to a con-  
 1166 tradiction:

1167 FACT 5.11. *For every  $i \in [r]$ , if  $u_i$  is the left child of  $v_i$  then*  
 1168  $|I_{u_i}| < p/3$ .

1169 PROOF. Indeed,  $u_i$  being at the left of the path to the leaf at  
 1170 position  $3kp$ ,  $I_{u_i} \subseteq [1, 3kp] \subseteq A$ . But  $\text{dist}(I_{u_i}, A) > d/2 - p/3$ , so  
 1171 it must be that  $|I_{u_i}| < p/3$ .  $\square$

1172 FACT 5.12. *For every  $i \in [r]$ , if  $u_i$  is the right child of  $v_i$  then*  
 1173  $|I_{u_i}| < 5p$ .

1174 PROOF. Likewise, we now have  $I_{u_i} \subseteq [3kp + 1, d]$ . Intuitively,  
 1175 a large interval in this zone must contain roughly twice as much  
 1176 elements from  $A^c$  than from  $A$ , so they cannot be at distance close  
 1177 to the maximum  $d/2$ .

1178 Let  $l$  be the number of blocks of the form  $[3(k+i)p + 2p +$   
 1179  $1, 3(k+i+1)p] \subseteq A$  which intersects  $I_{u_i}$ . By contradiction, assume  
 1180 that  $|I_{u_i}| > 5p$ . Note that it implies that  $l \geq 2$ .

1181 Assume that  $l = 2$ , then  $|A \cap I_{u_i}| \leq 2p$  hence, as  $|I_{u_i}| \geq 5p$ ,  
 1182  $|A^c \cap I_{u_i}| \geq 3p$ . Therefore, using Lemma 5.1,  $d(A, I_{u_i}) \leq d/2 - p \leq$   
 1183  $d_2 - p/3$ .

1184 Finally, assume that  $l > 2$ . Then  $|A \cap I_{u_i}| \leq pl$  and  $|A^c \cap I_{u_i}| \geq$   
 1185  $2p(l-1)$ . Therefore, using Lemma 5.1,  $d(A, I_{u_i}) \leq d/2 - pl + 2p \leq$   
 1186  $d/2 - p \leq d_2 - p/3$ .  $\square$

1187 FACT 5.13. *It must be that  $r \geq 7kp$ .*

1188 PROOF. Indeed, since  $[1, d] \setminus \{3kp\} = [1, 12kp] \setminus \{3kp\}$  is covered  
 1189 by the  $I_{u_i}$ , which have size bounded by  $5p$  (thanks to Fact 5.11  
 1190 and Fact 5.12) and among which all but  $k$  may have size  $> 1$  (as  
 1191 we consider a circuit of non-skew depth  $k$ ), there must be at least  
 1192  $12kp - 5kp = 7kp$  of them.  $\square$

1193 FACT 5.14. *There is some index  $i_0$  such that  $u_{i_0}, u_{i_0+1}, \dots, u_{i_0+20p/3-1}$*   
 1194 *are all leaves in  $s$ .*

1195 PROOF. Indeed, only  $k$  among the  $7kp$   $u_i$ 's may not be leaves.  
 1196 By contradiction assume that all blocks of consecutive leaves have

1197 length smaller than  $20p/3$ , so overall length is  $(20p/3)(k + 1) +$   
 1198  $k < 7kp$  as we initially assumed that  $k, p \geq 30$ . This contradicts  
 1199 Fact 5.13. □

1200 We now consider the increasing sequence of intervals  $I_{v_{i_0+20p/3-1}} \subseteq$   
 1201  $I_{v_{i_0+20p/3-2}} \subseteq \dots \subseteq I_{v_{i_0}}$  (where the nodes  $u_{i_0}, u_{i_0+1}, \dots, u_{i_0+20p/3-1}$   
 1202 are those given by Fact 5.14), which we simply denote  $I_1 \subseteq I_2 \subseteq$   
 1203  $\dots \subseteq I_{20p/3}$ . Each  $I_i = [a_i, b_i]$  contains  $3kp$ , and  $|I_{i+1}| = |I_i| +$   
 1204  $1$ . We let  $n_i = |I_i \cap A|$  and  $m_i = |I_i \cap A^c|$ . The assumption  
 1205  $d(A, I_i) > d/2 - p/3$  can be rephrased, thanks to Lemma 5.1, as  
 1206  $|n_i - m_i| \leq p/3$ .

1207 First, note that for all  $j < 6p$ ,  $b_j \notin \{3(k+i)p + 2p + 1 \mid 1 \leq i \leq$   
 1208  $3k\}$ . Indeed, for such a  $j$  one would have  $|n_{j+2p/3+1} - m_{j+2p/3+1}| =$   
 1209  $|n_j - m_j| + 2p/3 + 1 > p/3$  leading to a contradiction. Therefore,  
 1210 all the  $b_j$  for  $j = 1, \dots, 6p - 1$  belong to  $[3(k+i)p + 2p + 2, 3(k +$   
 1211  $i + 1)p + 2p]$  for some  $1 \leq i \leq 3k$ . Hence,  $m_{6p-1} - m_1 \leq 2p$ , which  
 1212 implies that  $n_{6p-1} - n_1 \geq 4p$ .

1213 Finally,

$$\begin{aligned} 2p/3 &\geq \left| |n_1 - m_1| - |n_{6p-1} - m_{6p-1}| \right| \\ &\geq |n_{6p-1} - m_{6p-1}| - |n_1 - m_1| \\ &\geq n_{6p-1} - m_{6p-1} + m_1 - n_1 \\ &\geq 4p - 2p \end{aligned}$$

1214 which leads to a contradiction and concludes the proof. □

1215 **5.1.2. Circuits with Many Parse Trees.** We now turn our  
 1216 focus to *k-PT circuits* which are circuits with at most  $k$  different  
 1217 parse trees. We first start by a key technical lemma that works  
 1218 both in the non-commutative and commutative (later discussed in  
 1219 Section 5.2) settings.

**Balanced Subsets** For  $s \in \text{Tree}_d$  and  $X \subseteq [d]$ , we define

$$\text{dist}(X, s) = \min \{ \text{dist}(X, A) \mid A \text{ spanned by } s \}.$$

1220 In the following, we let  $\binom{[d]}{d/2}$  denote the subsets of  $[d]$  of size  $d/2$ .  
 1221 For a subset  $\mathcal{P} \subseteq 2^{[d]}$  we write  $\mathcal{U}(\mathcal{P})$  for the uniform distribution  
 1222 over  $\mathcal{P}$ .

1223 Recall that, following Lemma 5.1, if  $X \in \binom{[d]}{d/2}$  and  $A \subseteq [d]$ ,  
 1224  $\text{dist}(X, A) > d/2 - \delta$  rewrites as  $||A \cap X| - |A^c \cap X|| \leq \delta$ , meaning  
 1225 that  $A$  is  $X$ -balanced.

1226 The following lemma is a subtle probabilistic analysis bounding  
 1227 the number of subsets that are balanced over all subsets spanned  
 1228 by a given fixed shape  $s$ . This will later entail the existence of a  
 1229 subset which is close to all parses trees in  $\text{PT}(\mathcal{C})$ , provided  $|\text{PT}(\mathcal{C})|$   
 1230 is not too large. It holds in both the non-commutative (in which  
 1231 it was originally proved) and the commutative settings.

1232 LEMMA 5.15 (Adapted from Claim 15 in Lagarde *et al.* 2018). *Let*  
 1233  *$s \in \text{Tree}_d$  be a shape with  $d$  leaves, and  $\delta \leq \sqrt{d}/2$ . Then*

$$\Pr_{X \sim \mathcal{U}\left(\binom{[d]}{d/2}\right)} [\text{dist}(X, s) > d/2 - \delta] \leq 2^{-\alpha d/\delta^2},$$

1234 *where  $\alpha$  is some positive constant.*

1235 We shall use an intermediate result from the aforementioned  
 1236 paper. Their proof (based on a greedy construction) can be read  
 1237 just as such in the commutative setting.

1238 LEMMA 5.16 (Subclaim 21 in Lagarde *et al.* 2018). *Let  $s \in \text{Tree}_d$ ,*  
 1239 *and  $r, t$  be integers such that  $rt \leq d/4$ . Then there exists a se-*  
 1240 *quence  $u_1, \dots, u_r$  of nodes of  $s$  such that for all  $i \in [r]$ ,*

$$\left| A_{v_i} \setminus \left( \bigcup_{j=1}^{i-1} A_{u_j} \right) \right| \geq t.$$

1241 We now give the proof of Lemma 5.16 in the commutative set-  
 1242 ting, noting that the proof in the non-commutative setting is a  
 1243 restricted version of the one we give, where spanned subsets of  $[d]$   
 1244 are replaced by spanned intervals.

1245 PROOF. We pick  $t = \delta^2$  and  $r = \frac{d}{4\delta^2}$ , and apply Lemma 5.16 to  
 1246 obtain a sequence  $v_1, \dots, v_r$  of nodes of  $s$ . Then we have:

$$\begin{aligned}
 & \Pr_{X \sim \mathcal{U}\left(\binom{[d]}{d/2}\right)} [\text{dist}(X, s) > d/2 - \delta] \\
 &= \Pr_{X \sim \mathcal{U}\left(\binom{[d]}{d/2}\right)} [\text{for all node } v \text{ of } s, ||X \cap A_v| - |X^c \cap A_v|| \leq \delta] \\
 &\leq d \Pr_{X \sim \mathcal{U}(2^{[d]})} [\text{for all node } v \text{ of } s, ||X \cap A_v| - |X^c \cap A_v|| \leq \delta]
 \end{aligned}$$

The last inequality follows from the general fact, applied using  $2^d \leq d \binom{d}{d/2}$ , that, for any event  $E$  and finite subsets  $\mathcal{P} \subseteq \mathcal{P}'$  with  $|\mathcal{P}'| \leq k|\mathcal{P}|$  one has

$$\Pr_{A \sim \mathcal{U}(\mathcal{P})}(E) \leq k \Pr_{A \sim \mathcal{U}(\mathcal{P}')} (E).$$

1247 Following from there, we let  $E_i$ , for  $i \in [r]$ , be the event  $||X \cap$   
 1248  $A_{v_i}| - |X^c \cap A_{v_i}|| \leq \delta$ , and obtain

$$\begin{aligned}
 & d \Pr_{X \sim \mathcal{U}(2^{[d]})} [\text{for all node } v \text{ of } s, ||X \cap A_v| - |X^c \cap A_v|| \leq \delta] \\
 &\leq d \Pr_{X \sim \mathcal{U}(2^{[d]})} [\forall i \in [r], E_i] \\
 &\leq d \prod_{i=1}^r \Pr_{X \sim \mathcal{U}(2^{[d]})} [E_i \mid \forall j < i, E_j]
 \end{aligned}$$

1249 In order to bound the terms  $\Pr_{X \sim \mathcal{U}(2^{[d]})}[E_i \mid \forall j < i, E_j]$  we use  
 1250 the following consequence of the Central Limit theorem.

1251 **FACT 5.17.** *There exist  $\beta < 1$  such that for all random variable  $Y$*   
 1252 *following an unbiased binomial law of parameter  $n$ , and all interval*  
 1253  *$I$  with  $|I| \leq 2\sqrt{n}$ , one has  $\Pr(Y \in I) \leq \beta$ .*

1254 If  $X$  is sampled uniformly among  $[d]$  and  $X \cap \left(\bigcup_{j < i} A_{u_j}\right)$  is  
 1255 fixed, let  $e = |X \cap \left(\bigcup_{j < i} A_{u_j}\right)| - |X^c \cap \left(\bigcup_{j < i} A_{u_j}\right)|$ . Then the  
 1256 event  $E_i$  can be rephrased as having a random variable following

1257 an unbiased binomial law of parameter  $t = \delta^2$  sit in  $[-\delta - e, \delta - e]$   
 1258 of size  $2\delta$ , which is bounded by  $\beta$  thanks to Fact 5.17. Hence,

$$\Pr_{X \sim \mathcal{U}\left(\binom{[d]}{d/2}\right)} [\text{dist}(X, s) > d/2 - \delta] w \leq d\beta^r = d\beta^{\frac{d}{4\delta^2}} \leq 2^{-\alpha d/\delta^2}$$

1259 for some positive constant  $\alpha$ . □

1260 **Superpolynomial lower bounds** Lagarde, Limaye, and Srinivasan (Lagarde *et al.* 2018) obtained a superpolynomial lower bound  
 1261 for superpolynomial  $k$  (up to  $k = 2^{d^{\frac{1}{3}-\epsilon}}$ ). In the statement below,  
 1262 the first item shows how to obtain the same result using Theorem  
 1263 4.1, while the second item improves the previous bound by  
 1264 applying Theorem 4.2 instead.  
 1265

1266 **THEOREM 5.18.** *Let  $f$  be a homogeneous non-commutative polynomial of degree  $d$  and with  $n$  variables such that, for all  $A \subseteq [d]$ ,*  
 1267  *$M_A(f)$  has full rank. Let  $\epsilon > 0$ . Then for large enough  $d$ ,*  
 1268

1269 (i) any  $2^{d^{1/3-\epsilon}}$ -PT circuit computing  $f$  has size at least  $2^{d^{1/3}(\log n - d^{-\epsilon})}$ ;

1270 (ii) any  $2^{d^{1-\epsilon}}$ -PT circuit computing  $f$  has size at least  $n^{d^{\epsilon/3}} d^{-2}$ .

PROOF. Let  $\mathcal{C}$  be a  $k$ -PT circuit computing  $f$ , and  $\delta \leq \sqrt{d}$ . We first show that there exists a subset  $A \subseteq [d]$  which is close to all parse trees in  $\mathcal{C}$ . Indeed, a union bound and Lemma 5.15 yield

$$\begin{aligned} & \Pr_{A \sim \mathcal{U}\left(\binom{[d]}{d/2}\right)} [\exists s \in \text{PT}(\mathcal{C}), \text{dist}(A, s) > d/2 - \delta] \\ & \leq \sum_{s \in \text{PT}(\mathcal{C})} \Pr_{A \sim \mathcal{U}\left(\binom{[d]}{d/2}\right)} [\text{dist}(A, s) > d/2 - \delta] \leq k2^{-\alpha d/\delta^2} \end{aligned}$$

1271 for large enough  $d$ .

1272 Choosing appropriate values for  $\delta$  and  $k$  and applying Theorem  
 1273 4.1 (*resp.* Theorem 4.2) leads the first (*resp.* second) item.

(i) Choosing  $\delta = d^{1/3}$  and  $k = 2^{d^{1/3-\epsilon}}$ , we have that  $k2^{-\alpha d/\delta^2} = 2^{d^{1/3-\epsilon} - \alpha d^{1/3}} < 1$ . This implies the existence of a subset  $A \subseteq [d]$  of size  $d/2$  such that for all  $s \in \text{PT}(\mathcal{C})$ ,  $\text{dist}(A, s) \leq d/2 -$

$\delta$ , that is, any  $s \in \text{PT}(\mathcal{C})$  spans an interval  $I(s)$  at distance at most  $d/2 - \delta$  from  $A$ . Finally, we apply Theorem 4.1 to obtain

$$|\mathcal{C}| \geq \text{rank}(M_A(f)) n^{-(d/2-\delta)} k^{-1} = n^{d/2} n^{-(d/2-d^{1/3})} 2^{-d^{1/3-\varepsilon}} = 2^{d^{1/3}(\log n - d^{-\varepsilon})}.$$

1274 (ii) Choosing  $\delta = d^{\varepsilon/3}$  and  $k = 2^{d^{1-\varepsilon}}$ , we have that  $k2^{-\alpha d/\delta^2} =$   
 1275  $2^{d^{1-\varepsilon} - \alpha d^{1-\frac{2}{3}\varepsilon}} < 1$ , which again lets us choose  $A \subseteq [d]$  of size  
 1276  $d/2$  and such that for all  $s \in \text{PT}(\mathcal{C})$ ,  $\text{dist}(s, A) \leq d/2 - \delta$ .  
 1277 Now, applying Theorem 4.2 we obtain

$$|\mathcal{C}| \geq \text{rank}(M_A(f)) n^{-(d/2-\delta)} d^{-2} = n^\delta d^{-2} = n^{d^{\varepsilon/3}} d^{-2}.$$

□

1278 In the second item, the bound  $2^{d^{1-\varepsilon}}$  on the number of parse  
 1279 trees is to be compared to the total number of shapes of size  $d$   
 1280 which is bounded by  $2^{2d}$  as noticed in Remark 3.2. As explained in  
 1281 the introduction this means that we obtain superpolynomial lower  
 1282 bounds for any class of circuits which has a small defect in the  
 1283 exponent of the total number of parse trees.

1284 **5.2. Applications in the commutative setting.** Regarding  
 1285 application in the commutative setting, we again consider the class  
 1286 of  $k$ -PT circuits which are set-multilinear circuits with at most  $k$   
 1287 different commutative parse trees. Recall from Section 4.3 that  
 1288 in the commutative set-multilinear setting, parse trees are shapes  
 1289 whose leaves are labelled by integers without repetition. In par-  
 1290 ticular the number of parse trees is roughly bounded by  $d!$  (see  
 1291 Remark 4.11).

1292 Arvind and Raja (Arvind & Raja 2016) showed a superpoly-  
 1293 nomial lower bound for  $k$ -PT circuits computing set-multilinear  
 1294 polynomial for sublinear  $k$  (up to  $k = d^{1/2-\varepsilon}$ ). We improve this  
 1295 to superpolynomial  $k$  (up to  $k = 2^{d^{1-\varepsilon}}$ ).

1296 However, the generic lower bound theorems, namely Theorem 4.8  
 1297 and Theorem 4.9, are not exactly the same, so we obtain two incom-  
 1298 parable bounds. In the following lower bounds, the set-multilinear  
 1299 polynomials that we consider have their variables partitionned into  
 1300  $X = X_1 \sqcup X_2 \sqcup \dots \sqcup X_d$  with  $n = |X_i|$  for all  $i$ .



1301 THEOREM 5.19. Let  $f$  be a set-multilinear commutative polynomial such that for all  $A \subseteq [d]$ , the matrix  $M_A(f)$  has full rank. Let  
 1302  $\varepsilon > 0$ . Then for large enough  $d$ ,

- 1304 (i) any  $2^{d^{1/3-\varepsilon}}$ -PT circuit computing  $f$  has size at least  $2^{d^{1/3}(\log n - d^{-\varepsilon})}$ ;  
 1305 (ii) any  $2^{d^{1-\varepsilon}}$ -PT circuit computing  $f$  has size at least  $n^{d^{\varepsilon/3}} d^{-2}$ .  
 1306 In particular, this lower bound is super polynomial when  $d$  is  
 1307 at most a polynomial in  $\log n$ .

PROOF. Let  $\mathcal{C}$  be a  $k$ -PT circuit computing  $f$ , and  $\delta \leq \sqrt{d}$ . By union bound and Lemma 5.15 for the commutative setting,

$$\begin{aligned} & \Pr_{A \sim \mathcal{U}\left(\binom{[d]}{d/2}\right)} \left[ \exists s \in \text{PT}(\mathcal{C}), \text{dist}(A, s) > d/2 - \delta \right] \\ & \leq \sum_{s \in \text{PT}(\mathcal{C})} \Pr_{A \sim \mathcal{U}\left(\binom{[d]}{d/2}\right)} \left[ \text{dist}(A, s) > d/2 - \delta \right] \leq k 2^{-\alpha d / \delta^2} \end{aligned}$$

1308 Choosing appropriate values for  $\delta$  and  $k$  and applying Theo-  
 1309 rem 4.8 (*resp.* Theorem 4.9) leads the first (*resp.* second) item.

- (i) Choosing  $\delta = d^{1/3}$  and  $k = 2^{d^{1/3-\varepsilon}}$ , we have that  $k 2^{-\alpha d / \delta^2} = 2^{d^{1/3-\varepsilon} - \alpha d^{1/3}} < 1$ . Hence, picking a subset  $A \subseteq [d]$  of size  $d/2$  such that any  $s \in \text{PT}(\mathcal{C})$  spans an interval  $I(s)$  at distance at most  $d/2 - \delta$  from  $A$ , and applying Theorem 4.8 yields

$$\begin{aligned} |\mathcal{C}| & \geq \text{rank}(M_A(f)) n^{-(d/2-\delta)} k^{-1} = n^{d/2} n^{-(d/2-d^{1/3})} 2^{-d^{1/3-\varepsilon}} \\ & = 2^{d^{1/3}(\log n - d^{-\varepsilon})}. \end{aligned}$$

- 1310 (ii) Choosing  $\delta = d^{\varepsilon/3}$  and  $k = 2^{d^{1-\varepsilon}}$ , we have that  $k 2^{-\alpha d / \delta^2} =$   
 1311  $2^{d^{1-\varepsilon} - \alpha d^{1-\frac{2}{3}\varepsilon}} < 1$ . Hence, picking a subset  $A \subseteq [d]$  of size  $d/2$   
 1312 and such that for all  $s \in \text{PT}(\mathcal{C})$ ,  $\text{dist}(s, A) \leq d/2 - \delta$ , and  
 1313 applying Theorem 4.9 yields

$$|\mathcal{C}| \geq \text{rank}(M_A(f)) n^{-(d/2-\delta)} k^{-1} = n^\delta 2^{-d^{1-\varepsilon}} = n^{d^{\varepsilon/3}} 2^{-d}. \quad \square$$

1314

## 6. Discussion

1315 We presented a new tool for proving lower bounds for arithmetic  
 1316 circuits in the form of the Hankel matrix. We obtained strong  
 1317 lower bounds both in the commutative and non-commutative set-  
 1318 tings using generic decompositions of the Hankel matrix. A natural  
 1319 question is how far this approach can be pushed. The first remark  
 1320 is that the rank of the Hankel matrix is exactly the size of the  
 1321 smallest circuit computing a given (non-associative) polynomial,  
 1322 hence the potential loss can only be in analyzing the Hankel ma-  
 1323 trix. Limaye, Malod and Srinivasan (Limaye *et al.* 2016) defined  
 1324 a polynomial computed by a circuit of polynomial size but such  
 1325 that all partial derivative matrices have full rank: this shows that  
 1326 one cannot use our decomposition of the Hankel matrix to obtain  
 1327 strong lower bounds for the class of all circuits. This limitation is  
 1328 an invitation to get a deeper understanding of the Hankel matrix  
 1329 and to find other ways of decomposing it.

1330 On a different perspective, the Hankel matrix has been suc-  
 1331 cessfully used as a data structure for learning algorithms (in both  
 1332 supervised and unsupervised settings). It is tempting, using the  
 1333 characterization that we present in this paper, to construct algo-  
 1334 rithms for learning polynomials relying on the Hankel matrix as  
 1335 algorithmic representation.

1336

## References

1337 MANINDRA AGRAWAL, ROHIT GURJAR, ARPITA KORWAR & NITIN  
 1338 SAXENA (2015). Hitting-Sets for ROABP and Sum of Set-Multilinear  
 1339 Circuits. *SIAM Journal on Computing* **44**(3), 669–697.

1340 ERIC ALLENDER, JIA JIAO, MEENA MAHAJAN & V. VINAY (1998).  
 1341 Non-Commutative Arithmetic Circuits: Depth Reduction and Size  
 1342 Lower Bounds. *Theoretical Computer Science* **209**(1-2), 47–86.

1343 VIKRAMAN ARVIND, RAJIT DATTA, PARTHA MUKHOPADHYAY &  
 1344 S. RAJA (2017). Efficient Identity Testing and Polynomial Factorization  
 1345 in Nonassociative Free Rings. In *Proceedings of the 42nd International  
 1346 Symposium on Mathematical Foundations of Computer Science (MFCS  
 1347 2017)*, volume 83 of *LIPICs*, 38:1–38:13. Schloss Dagstuhl - Leibniz-  
 1348 Zentrum fuer Informatik. ISBN 978-3-95977-046-0. ISSN 1868-8969.

- 1349 VIKRAMAN ARVIND & S. RAJA (2016). Some Lower Bound Results for  
1350 Set-Multilinear Arithmetic Computations. *Chicago Journal of Theoret-*  
1351 *ical Computer Science* **2016**.
- 1352 WALTER BAUR & VOLKER STRASSEN (1983). The Complexity of Par-  
1353 tial Derivatives. *Theoretical Computer Science* **22**, 317–330.
- 1354 SYMEON BOZAPALIDIS & OLYMPIA LOUSCOU-BOZAPALIDOU (1983).  
1355 The Rank of a Formal Tree Power Series. *Theoretical Computer Science*  
1356 **27**, 211–215.
- 1357 MARCO L. CARMOSINO, RUSSELL IMPAGLIAZZO, SHACHAR LOVETT &  
1358 IVAN MIHAJLIN (2018). Hardness Amplification for Non-Commutative  
1359 Arithmetic Circuits. In *Proceedings of the 33rd Computational Complex-*  
1360 *ity Conference (CCC 2018)*, volume 102 of *LIPICs*, 12:1–12:16. Schloss  
1361 Dagstuhl - Leibniz-Zentrum fuer Informatik.
- 1362 SURYAJITH CHILLARA & PARTHA MUKHOPADHYAY (2019). Depth-4  
1363 Lower Bounds, Determinantal Complexity: A Unified Approach. *Com-*  
1364 *putational Complexity* **28**(4), 545–572.
- 1365 ZEEV DVIR, GUILLAUME MALOD, SYLVAIN PERIFEL & AMIR YEHU-  
1366 DAYOFF (2012). Separating multilinear branching programs and for-  
1367 mulas. In *Proceedings of the 44th Symposium on Theory of Computing*  
1368 *Conference (STOC 2012)*, 615–624. ACM.
- 1369 NATHANAËL FIJALKOW, GUILLAUME LAGARDE & PIERRE OHLMANN  
1370 (2018). Tight Bounds using Hankel Matrix for Arithmetic Circuits with  
1371 Unique Parse Trees. *Electronic Colloquium on Computational Com-*  
1372 *plexity (ECCC)* **25**, 38. URL [https://eccc.weizmann.ac.il/report/](https://eccc.weizmann.ac.il/report/2018/038)  
1373 [2018/038](https://eccc.weizmann.ac.il/report/2018/038).
- 1374 MICHEL FLIESS (1974). Matrices de Hankel. *Journal de Mathématiques*  
1375 *Pures et Appliquées* **53**, 197–222.
- 1376 MICHAEL A. FORBES, RAMPRASAD SAPTHARISHI & AMIR SHPILKA  
1377 (2014). Hitting sets for multilinear read-once algebraic branching pro-  
1378 grams, in any order. In *Proceedings of the 46th Symposium on Theory*  
1379 *of Computing, (STOC 2014)*, 867–875. ACM.
- 1380 HERVÉ FOURNIER, NUTAN LIMAYE, GUILLAUME MALOD & SRIKANTH  
1381 SRINIVASAN (2014). Lower bounds for depth 4 formulas computing

- 1382 iterated matrix multiplication. In *Proceedings of the 46th Symposium*  
1383 *on Theory of Computing (STOC 2014)*, 128–135. ACM.
- 1384 ANKIT GUPTA, PRITISH KAMATH, NEERAJ KAYAL & RAMPRASAD  
1385 SAPTHARISHI (2014). Approaching the Chasm at Depth Four. *J. ACM*  
1386 **61**(6), 33:1–33:16. URL <https://doi.org/10.1145/2629541>.
- 1387 ROHIT GURJAR, ARPITA KORWAR, NITIN SAXENA & THOMAS THIER-  
1388 AUF (2017). Deterministic Identity Testing for Sum of Read-Once Obliv-  
1389 ious Arithmetic Branching Programs. *Computational Complexity* **26**(4),  
1390 835–880.
- 1391 PAVEL HRUBEŠ, AVI WIGDERSON & AMIR YEHUDAYOFF (2010). Re-  
1392 lationless Completeness and Separations. In *Proceedings of the 25th*  
1393 *Annual IEEE Conference on Computational Complexity (CCC 2010)*,  
1394 280–290. IEEE Computer Society.
- 1395 PAVEL HRUBEŠ, AVI WIGDERSON & AMIR YEHUDAYOFF (2011). Non-  
1396 commutative circuits and the sum-of-squares problem. *Journal of the*  
1397 *American Mathematical Society* **24**(3), 871–898.
- 1398 LAURENT HYAFIL (1977). The Power of Commutativity. In *Proceedings*  
1399 *of the 18th Annual Symposium on Foundations of Computer Science*  
1400 *(FOCS 1977)*, 171–174. IEEE Computer Society.
- 1401 MARK JERRUM & MARC SNIR (1982). Some Exact Complexity Results  
1402 for Straight-Line Computations over Semirings. *J. ACM* **29**(3), 874–  
1403 897. URL <https://doi.org/10.1145/322326.322341>.
- 1404 VALENTINE KABANETS & RUSSELL IMPAGLIAZZO (2003). Derandom-  
1405 izing polynomial identity tests means proving circuit lower bounds. In  
1406 *Proceedings of the 35th Annual ACM Symposium on Theory of Comput-*  
1407 *ing (STOC 2003)*, 355–364. ACM.
- 1408 NEERAJ KAYAL, NUTAN LIMAYE, CHANDAN SAHA & SRIKANTH SRINI-  
1409 VASAN (2014a). An Exponential Lower Bound for Homogeneous Depth  
1410 Four Arithmetic Formulas. In *55th IEEE Annual Symposium on Foun-*  
1411 *dations of Computer Science, FOCS 2014, Philadelphia, PA, USA,*  
1412 *October 18-21, 2014*, 61–70. IEEE Computer Society. URL <https://doi.org/10.1109/FOCS.2014.15>.  
1413

- 1414 NEERAJ KAYAL, CHANDAN SAHA & RAMPRASAD SAPTHARISHI  
1415 (2014b). A super-polynomial lower bound for regular arithmetic for-  
1416 mulas. In *Proceedings of the 46th Symposium on Theory of Computing,*  
1417 *(STOC 2014)*, 146–153. ACM.
- 1418 MRINAL KUMAR & SHUBHANGI SARAF (2017). On the Power of Homo-  
1419 geneous Depth 4 Arithmetic Circuits. *SIAM J. Comput.* **46**(1), 336–387.  
1420 URL <https://doi.org/10.1137/140999335>.
- 1421 GUILLAUME LAGARDE, NUTAN LIMAYE & SRIKANTH SRINIVASAN  
1422 (2018). Lower Bounds and PIT for Non-commutative Arithmetic Cir-  
1423 cuits with Restricted Parse Trees. *Computational Complexity* 1–72.
- 1424 GUILLAUME LAGARDE, GUILLAUME MALOD & SYLVAIN PERIFEL  
1425 (2016). Non-commutative computations: lower bounds and polynomial  
1426 identity testing. *Electronic Colloquium on Computational Complexity*  
1427 *(ECCC)* **23**, 94.
- 1428 NUTAN LIMAYE, GUILLAUME MALOD & SRIKANTH SRINIVASAN  
1429 (2016). Lower Bounds for Non-Commutative Skew Circuits. *Theory*  
1430 *of Computing* **12**(1), 1–38.
- 1431 GUILLAUME MALOD & NATACHA PORTIER (2008). Characterizing  
1432 Valiant’s algebraic complexity classes. *Journal of Complexity* **24**(1),  
1433 16–38.
- 1434 NOAM NISAN (1991). Lower Bounds for Non-Commutative Computa-  
1435 tion (Extended Abstract). In *Proceedings of the 23rd Symposium on*  
1436 *Theory of Computing (STOC 1991)*, 410–418. ACM.
- 1437 NOAM NISAN & AVI WIGDERSON (1994). Hardness vs Randomness.  
1438 *Journal of Computer and System Sciences* **49**(2), 149–167.
- 1439 NOAM NISAN & AVI WIGDERSON (1997). Lower Bounds on Arithmetic  
1440 Circuits Via Partial Derivatives. *Computational Complexity* **6**(3), 217–  
1441 234.
- 1442 C. RAMYA & B. V. RAGHAVENDRA RAO (2018). Lower Bounds  
1443 for Special Cases of Syntactic Multilinear ABPs. In *Proceedings of*  
1444 *the 24th International Computing and Combinatorics Conference (CO-*  
1445 *COON 2018)*, volume 10976 of *Lecture Notes in Computer Science*, 701–  
1446 712. Springer.

1447 RAN RAZ & AMIR SHPILKA (2005). Deterministic polynomial identity  
 1448 testing in non-commutative models. *Computational Complexity* **14**(1),  
 1449 1–19.

1450 RAMPRASAD SAPTHARISHI & ANAMAY TENGSE (2017). Quasi-  
 1451 polynomial Hitting Sets for Circuits with Restricted Parse Trees. *Elec-  
 1452 tronic Colloquium on Computational Complexity (ECCC)* **24**, 135.

1453 N. J. A. SLOANE (editor) (2011). *The On-Line Encyclopedia of Integer  
 1454 Sequences*. Number of labeled rooted unordered binary trees (each node  
 1455 has out-degree  $\leq 2$ ), <https://oeis.org/A036774>.

1456 SEINOSUKE TODA (1992). Classes of Arithmetic Circuits Capturing the  
 1457 Complexity of Computing the Determinant. *IEICE Trans. Inf. Systems*  
 1458 **E75-D**(1), 116–124.

1459 LESLIE G. VALIANT (1979). The Complexity of Computing the Perma-  
 1460 nent. *Theoretical Computer Science* **8**, 189–201.

1461 Manuscript received

NATHANAËL FIJALKOW  
 CNRS, LaBRI, Bordeaux, France  
 The Alan Turing Institute of data  
 science, London, United King-  
 1462 dom  
 Nathanael.Fijalkow@labri.fr

GUILLAUME LAGARDE  
 LaBRI, Bordeaux, France  
 Guillaume.Lagarde@labri.fr

PIERRE OHLMANN  
 Université de Paris, IRIF, CNRS,  
 F-75013 Paris, France  
 1463 Pierre.Ohlmann@irif.fr

OLIVIER SERRE  
 Université de Paris, IRIF, CNRS,  
 F-75013 Paris, France  
 Olivier.Serre@cnrs.fr