Matrix identities are hard: Fast blackbox PIT for noncommutative exponential-size constant-depth homogeneous circuits

Foram Lakhani [*] Nitin Saxena [†]

Abstract

In this work we give a randomized blackbox polynomial identity testing (PIT) algorithm for constant-depth homogeneous noncommutative circuits, with poly-logarithmic time complexity in the circuit size. In fact, we show that the polynomial computable by such a depth-$(\Delta - 1)$ circuit of size s cannot be a polynomial identity for the $O(\log^{\Delta} s)$-dimension matrix algebra $\mathcal{M}_{O(\log^{\Delta} s)}(\mathbb{F})$, for a sufficiently large field $\mathbb{F}$.

This represents progress toward answering the question raised in [BW05, AJMR19] to find efficient randomized blackbox PIT algorithm for circuits that output polynomials with *doubly* exponential sparsity. [AJMR19, SBR25] defined constant-depth $+$-regular circuit and gave a PIT for it. Their model of size S is simulated by our model of size $\exp(S)$; hence, we recover their result. Because of the regularity condition their model is highly restrictive than ours.

We decompose a depth-Δ formula into its constituent depth-$(\Delta - 1)$ subformulas and analyze them concurrently using Hadamard algebra, subsequently translating these findings back to the depth-Δ setting. We reduce the multivariate case to the bivariate case involving the noncommuting variables x and y. Finally, we apply a transformation to x and y to ensure that the nonzero witness coefficient is associated with a monomial of exponentially-lower y-degree. We call this the phenomenon of *noncommutative low-degree concentration*.

Contents

[*] CSE, IIT Kanpur, foram@cse.iitk.ac.in
[†] CSE, IIT Kanpur, nitin@cse.iitk.ac.in

1 Introduction

Polynomial Identity Testing (PIT) is the problem of deciding whether a polynomial computed by an algebraic circuit is identically zero. Problems of practical significance like primality testing [AKS04] and parallel algorithms for finding perfect matchings reduce to PIT [CGG+26, FGT19]. Moreover, derandomizing blackbox PIT for depth-4 formulas has implications for VP $\neq$ VNP question which is an algebraic analogue of P $\neq$ NP [Val79, SV85]. This problem has been extensively studied under many restrictions (see the surveys [Sax14, Sax09]). Hyafil [Hya77] and Nisan [Nis91] imposed an algebraic restriction of noncommutativity to bound the complexity of computing the determinant polynomial.

The parallel between randomized PIT for commutative and noncommutative cases is that, in the commutative scenario, random field elements are substituted to test for nonzeroness, whereas in the noncommutative context, the polynomial is evaluated using random matrices [BW05]. An upper bound on the dimension of matrices is given by the classical Amitsur-Levitski result [AL50]. It shows that a noncommutative nonzero polynomial f of degree $2d-1$ does not vanish on the matrix algebra $\mathcal{M}_d(\mathbb{F}) \cong (\mathbb{F}^{d\times d}, +, *)$. Hence substitution of noncommutative variables by random matrices from $\mathbb{F}^{d\times d}$ is highly likely to be nonzero (assuming large $\mathbb{F}$). A general noncommutative circuit can have degree $2^{O(s)}$ and this approach yields a randomized *exponential*-time algorithm. Thus, unlike commutative PIT, noncommutative PIT is a practically hard problem currently.

There are interesting families of identities known on the matrix algebra, with connections to graph theory and cohomology theory [Kos09]. More broadly, the theory of polynomial identities studies algebraic structures through the polynomial equations they satisfy, illuminating universal laws governing their representations. A central question in this area is to characterize algebraic systems by the identities they satisfy. Matrix algebra is ubiquitous in mathematics and its applications, making the study of their polynomial identities well motivated from both theoretical and practical perspectives. [AL50] is the first step in this direction as it shows that matrix algebra $\mathcal{M}_d(\mathbb{F})$ admits no identity of degree less than $2d$. However, it is completely oblivious to the complexity of these identities. We conjecture that the identities are computationally hard. Some evidence was provided in [BW05, Sec.6.2], by studying the standard identity on matrices.

1.1 Prior work

The authors in [AJMR19] gave an efficient randomized algorithm for the PIT problem when the polynomial is of exponential degree, but has sparsity that is (only) exponential in the size of the circuit. General circuits of size s can compute polynomials that are doubly exponentially dense e.g. $(x+y)^{2^s}$. In [AJMR19] it has been shown that if the sparsity of the polynomial is t, then identity testing can be performed using matrices of dimension $O(\log t)$. To handle doubly-exponential sparsity for the special case of homogeneous noncommutative circuits, called $+$-regular circuits, they gave an efficient deterministic white-box PIT algorithm. For the blackbox setting, they obtain an efficient randomized PIT algorithm only for depth-3 $+$-regular circuits. In particular, they show that if a nonzero noncommutative polynomial $f \in \mathbb{F}\langle X\rangle$ is computed by a depth-3 $+$-regular circuit of size s, then f cannot be a polynomial identity for the matrix algebra $\mathcal{M}_s(\mathbb{F})$ for a sufficiently large field $\mathbb{F}$. They asked if the blackbox algorithm for depth-3 can be extended for constant depths. This was answered by [SBR25] by showing that a nonzero noncommutative polynomial computed by a depth-c $+$-regular circuit of size s cannot be a polynomial identity for the matrix algebra $\mathcal{M}_n(\mathbb{F})$ for $n = s^{O(c^2)}$.

However, as noted in [AJMR19], the computational power of $+$-regular circuits is limited. Nisan's rank based argument [Nis91] can be adapted to show that the noncommutative permanent cannot be computed by polynomial-size $+$-regular circuits.

The derandomization of noncommutative PIT is another open problem. When the circuit is restricted to be a formula the deterministic polynomial time white-box algorithms [GGOW16, IQS18, HH21], randomized polynomial time blackbox algorithm [DM17], and quasi-polynomial time blackbox algorithms [ACM24] are known. The question for general circuits remain open.

1.2 Our results

We give an efficient randomized blackbox algorithm for polynomials that are represented as constant-depth homogeneous noncommutative circuits of exponential-size. An algebraic circuit over a field $\mathbb{F}$ is a layered acyclic directed graph with a sink node called output node. The source nodes are called input nodes and are labeled by variables or field constants. The non-input nodes are labeled by $\times$ (multiplication gate) and $+$ (addition gate) in alternate layers. The edges may be labeled by field constants. The computation is defined in a natural way. The size of the circuit is the number of edges and vertices. The depth is the number of layers. The product-depth is the number of product layers.

Our algorithm solves PIT for polynomials that can have doubly exponential sparsity. Our model strictly generalizes constant-depth $+$-regular circuits. A depth-Δ $+$-regular circuit is a homogeneous circuit with Δ-layers of addition gates, where all gates in an addition layer have the same syntactic degree, no directed path exist between the addition gates of the same layer, and all input to output paths go through exactly one $+$-gate in each layer. By collapsing intermediate multiplication layers, a size-s, depth-Δ $+$-regular circuit can be transformed into a homogeneous circuit of depth-Δ and size-$\exp(s)$, allowing us to recover their PIT algorithm. Because enforcing a uniform syntactic degree across each addition layer is highly restrictive, our model gives PIT for a strictly larger class of polynomials; for instance, the polynomial $\sum_{i=1}^{i^2 \le 2^s} (x^i + x^{i-1}y + y^i)^i \cdot (x+y)^{2^s - i^2}$ cannot be computed by constant-depth $+$-regular circuit of polynomial size but can be efficiently computed with exponential size in our model.

The idea of concentration via shifts used in this paper is inspired from [ASS13] where it was first studied for constant depth commutative models. However those methods do not directly translate to results for noncommutative model of exponential size in a trivial way. Consequently, the maps used here for concentration are significantly more complex and require new proofs.

In this paper we prove the following results.

Theorem 1 (Blackbox-PIT)**.** *Any nonzero noncommutative homogeneous circuit of size s and depth $\Delta - 1$ is nonzero on $\mathcal{M}_{O(\log^{\Delta} s)}(L)$, where L is degree-$O(\log s)$ extension of $\mathbb{F}$. Thus for this class, randomized blackbox PIT is* $\mathrm{poly}(\log^{\Delta} s)$*-time.*

Remark 1.1. This gives us the first fast blackbox PIT algorithm for exponential-size constant-depth homogeneous noncommutative circuits.

Theorem 2 (Hardness)**.** *Every identity for the matrix algebra $\mathcal{M}_d(\mathbb{F})$ requires size* $\exp(\Omega(d^{1/\Delta}))$ *depth $\Delta - 1$ homogeneous noncommutative circuits.*

Remark 1.2. This gives us the first noncommutative lowerbound for all matrix identities; including the *standard identity* and *identity of algebraicity* that are known to exist for $\mathcal{M}_d(\mathbb{F})$ [BW05].

1.3 Proof outline

We begin by showing that the multivariate noncommutative PIT problem can be reduced to a bivariate noncommutative PIT problem with only a constant factor increase in circuit size and depth. We then apply a bivariate transformation ensuring that if f is nonzero, the transformed polynomial has a nonzero coefficient witness among its low y-degree monomials.

Our approach is inspired by the rank concentration techniques of [ASS13], which established a quasi-polynomial time blackbox PIT algorithm for constant-depth commutative formulas whose $\prod$-layers respect a variable partition. They analyzed formulas over a Hadamard algebra and demonstrated that applying a variable 'shift', $f(x+c)$, yields low-support monomials with nonzero coefficients whenever $f \neq 0$. Here, a low-support monomial contains at most $O(\log s)$ distinct variables, where s denotes the size of the depth-3 circuit computing f.

Because the depth-3 circuits in [ASS13] compute multilinear polynomials, an independent shift for each linear term within a product is possible. In our setting, however, we make no assumptions regarding disjoint variables for the polynomials feeding into a product gate. In fact, we reduce the multivariate case to a bivariate model where x and y are noncommuting indeterminates. Furthermore, their approach utilizes a transfer matrix that does not generalize to higher depths under our framework.

Low-degree Concentration Idea. Paralleling the low-support rank-concentration strategy of [ASS13], our goal is to ensure that the transformed polynomial has a nonzero monomial with y-degree at most ℓ, where ℓ is exponentially small with respect to the circuit size. To understand the technique for such a concentration, first consider a homogeneous $\sum\prod\sum$ circuit computing a polynomial $f \in \mathbb{F}\langle x, y\rangle$. f can be expressed in the form $f = \mathbf{1}^{\mathbf{T}} \cdot g_1 \star \ldots \star g_n$, where $\star$ is Hadamard product, $\cdot$ is matrix product and each g_i is a (homogeneous) linear polynomial over the vector space $\mathbb{F}^{\mathrm{poly}(s)}$.

We devise a corresponding commutative model $f^{\mathrm{com}} \in \mathbb{F}[\mathbf{X}, \mathbf{Y}]$ with the property that $f \equiv 0 \iff f^{\mathrm{com}} \equiv 0$. f^{com} is also expressible as $f^{\mathrm{com}} = \mathbf{1}^{\mathbf{T}} \cdot g_1^{\mathrm{com}} \star \ldots \star g_n^{\mathrm{com}}$. Inspired by the idea of shifts, we define a commutative map called the ℓ- subset cover map, φ_ℓ such that $\varphi_\ell(g_1^{\mathrm{com}} \star \ldots \star g_n^{\mathrm{com}})$ is ℓ-concentrated: any coefficient corresponding to a monomial with $\mathbf{Y}$-degree greater than or equal to ℓ is in the span of coefficients that correspond to monomials with $\mathbf{Y}$-degree less than ℓ. Here, to derive the dependency, we use a new technique of partial derivative instead of transfer matrix of [ASS13]. ℓ-concentration of $\varphi_\ell(g_1^{\mathrm{com}} \star \ldots \star g_n^{\mathrm{com}})$ implies ℓ-concentration of $\varphi_\ell(f^{\mathrm{com}})$.

The noncommutative transformation, ψ_ℓ acts on f by tensoring x and y with matrices of dimension $\ell = O(\log s)$. Specifically, $x \mapsto A \otimes x$ and $y \mapsto B \otimes (x + y)$, where $A = (a_{ij})$, $B = (b_{ij})$ and a_{ij}, b_{ij} are commuting indeterminates. Under this map, the coefficients are matrices with entries in the polynomial ring $\mathbb{F}[a_{ij}, b_{ij} \mid i, j]$.

The $(1, \ell)$-th entry of a coefficient in $\psi_\ell(f)$ has a correspondence with $\varphi_\ell(f^{\mathrm{com}})$. The ℓ-concentration of commutative model $\varphi_\ell(f^{\mathrm{com}})$ together with this correspondence can be used to derive: if a coefficient of a higher y-degree monomial is nonzero, then a coefficient of a lower y-degree monomial must also be nonzero in $\psi_\ell(f)$. Crucially, this property is achieved using matrices A and B of dimension merely $O(\log s)$, which is the non-trivial part.

After concentration the coefficients of low-degree monomials of $\psi_\ell(f)$ are isolated when we substitute x and y with generic matrices M_x and M_y of dimension ℓ, whose entries are distinct commuting indeterminates. Following the reasoning in [AJMR19], it can be shown that monomial with a y-degree strictly less than ℓ generates a unique commutative monomial in the corresponding matrix product. The entries of the matrix $f(M_x, M_y)$ are commutative polynomials, at least one of which inherits these isolated coefficients and is thus nonzero.

Consequently, the combined map given by $x \mapsto A \otimes M_x$ and $y \mapsto B \otimes (M_x + M_y)$ effectively substitutes x and y with symbolic matrices of dimension $O(\log^2 s)$. The nonzero entry of the resulting matrix is a commutative polynomial and the Polynomial Identity Lemma yields a randomized PIT over an extension field of degree-$O(\log s)$.

1.3.1 Product depth-2 case

A product depth-2 circuit computing f can be expressed as $f = \mathbf{1}^{\mathbf{T}} \cdot h_1 \star \ldots \star h_n$, where $\star$ is Hadamard product, $\cdot$ is matrix product and each h_i is a product depth-1 circuit. The proof follows the same template. First we prove concentration in the f^{com} after applying the appropriate subset cover map, φ_{ℓ_2}. Then establish a connection between $\psi_{\ell_2}(f)$ and the commutative model. Using the concentration in commutative model, argue for the nonzeroness in low y-degree coefficients in $\psi_{\ell_2}(f)$. Finally, reduce the check of nonzeroness of these fewer coefficients to a commutative PIT check by substituting x, y with matrices.

The difference in this case is that h_i's are now polynomials and not linear forms. Using their concentration properties from the previous case, the analogous results are derived for $\ell_2 = O(\log^2 s)$. The composition of these steps yields the transformation $x \mapsto A \otimes M_x, y \mapsto B \otimes (M_x + M_y)$ which substitutes x and y with matrices of cumulative dimension $O(\log^4 s)$. This framework directly scales via induction to any constant depth.

2 Basics

Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. We denote a size-ℓ subset of $[n]$ by $S \in \binom{[n]}{\ell}$. The matrix product is denoted by $\cdot$ and the Hadamard product (co-ordinate wise product) by $\star$. For a noncommuting monomial m, $m(i)$ is the variable at the index i when m is interpreted as a string. Define $\deg_y(m) := |\{i \mid m(i) = y\}|$ when m is a noncommuting monomial. For a monomial m on commutative variables $\mathbf{X} = \{X_1, ..., X_n\}, \mathbf{Y} = \{Y_1, ..., Y_n\}$, $\mathbf{Y}$-degree of m is $\sum_{Y_i \in \mathbf{Y}} \deg_{Y_i}(m) =: \deg_{\mathbf{Y}}(m)$. For any commutative monomial m, let $\mathcal{Y}(m) := \{i : Y_i \mid m\}$.

Hadamard algebra. Let $\mathbb{F}$ be a commutative ring and let $k \in \mathbb{N}$. The Hadamard algebra is defined as $H_k(\mathbb{F}) = (\mathbb{F}^k, +, \star)$, where addition and multiplication are performed coordinate-wise. $H_k(\mathbb{F})$ is a commutative $\mathbb{F}$-algebra with identity element $\bar{1}$ and additive identity $\bar{0}$. The polynomial ring $H_k(\mathbb{F})[x]$ is collection of all polynomials with coefficients in $H_k(\mathbb{F})$. Addition and multiplication of $H_k(\mathbb{F})$ naturally extend to this ring. There is a natural correspondence between $H_k(\mathbb{F})[x]$ and $H_k(\mathbb{F}[x])$ obtained by viewing elements as vectors whose entries are polynomials over $\mathbb{F}$. A unit in $H_k(\mathbb{F})$ or $H_k(\mathbb{F}(x))$ is the one that has all nonzero co-ordinates and its inverse is the coordinate-wise inverse.

Recall the Polynomial Identity Lemma (commutative setting), which is widely known as the Ore-DeMillo-Lipton-Schwartz-Zippel Lemma:

Lemma 2.1. *[Ore22, DL78, Zip79, Sch80] [The Polynomial Identity Lemma] Let $f(x_1, x_2, \ldots, x_n)$ be a nonzero, degree-d n-variate polynomial over a field $\mathbb{F}$, and let $S \subset \mathbb{F}$ be a finite subset. Then, the number of zeros of f contained in the cartesian product S^n is bounded by $d \cdot |S|^{n-1}$.*

The commutative transformation (n-subset cover map) φ_n. Let $\mathbf{a} := \{a_{ij} \mid i, j \in [n]\}, \mathbf{b} := \{b_{ij} \mid i, j \in [n]\}$ and d be the degree of the polynomial to which this map is applied.

We define $\varphi_n(X_i) = \eta_i(\alpha, \mathbf{a})X_i$ and $\varphi_n(Y_i) = \rho_i(\alpha, \mathbf{b})(X_i + Y_i)$ where η_i, ρ_i are interpolating polynomials defined as follows.

Let a set $S = (s_1, \ldots, s_t) \subseteq [d]$ of size at most n (i.e., $t \leq n$), with $s_0 := 0 < s_1 < s_2 < \cdots < s_t < s_{t+1} := d + 1$ uniquely map to an element $\alpha_S \in \mathbb{F}$. We define η_i and ρ_i for all $i \in [d]$, as interpolating polynomials such that

$$\rho_i(\alpha_S, \mathbf{b}) = \begin{cases} b_{k,k} & \text{if } s_{k-1} < i < s_k \\ b_{k,k+1} & \text{if } i = s_k \end{cases} \tag{2.2}$$

$$\eta_i(\alpha_S, \mathbf{a}) = \begin{cases} a_{k,k} & \text{if } s_{k-1} < i < s_k \\ a_{k,k+1} & \text{if } i = s_k \end{cases} \tag{2.3}$$

Construction shows that after setting α to α_S, for all $i \in S$ there is a $j \in [n]$ unique to this i, such that $\rho_i(\alpha_S, \mathbf{b}) = b_{j,j+1}$ and $\eta_i(\alpha_S, \mathbf{b}) = a_{j,j+1}$. We will denote by $\varphi_{n,S}$ the specialized transformation $X_i \mapsto \eta_i(\alpha_S, \mathbf{a})X_i$ and $Y_i \mapsto \rho_i(\alpha_S, \mathbf{b})(X_i + Y_i)$. In contexts where n is fixed, we will drop the subscript n and denote φ_n by φ and $\varphi_{n,S}$ by φ_S.

The noncommutative transformation ψ_n. We denote a $(n+1)\times(n+1)$ matrix with commuting variables in diagonal and super-diagonal entries and zero elsewhere by

$$M_{\mathbf{a},n} = \begin{bmatrix} a_{1,1} & a_{1,2} & 0 & \dots & 0 & 0 \\ 0 & a_{2,2} & a_{2,3} & \dots & 0 & 0 \\ 0 & 0 & a_{3,3} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{n,n} & a_{n,n+1} \\ 0 & 0 & 0 & \dots & 0 & a_{n+1,n+1} \\ . \end{bmatrix}$$

Let $A := M_{\mathbf{a},n}, B := M_{\mathbf{b},n}$, a similar matrix as above but with $b_{i,j}$ instead of $a_{i,j}$. Define ψ_n as,

$$\begin{aligned} x &\mapsto A \otimes x \\ y &\mapsto B \otimes (x + y). \end{aligned}$$

For a polynomial over vectors with Hadamard multiplication and constant matrix product, ψ_n is defined so that it is a homomorphism.

For $u, v \in \mathbb{F}^w$, $\psi_n(ux) = u \otimes A \otimes x, \psi_n(vy) = v \otimes B \otimes x + v \otimes B \otimes y$. $\psi_n(f + g) = \psi_n(f) + \psi_n(g)$ $\psi_n(f \star g) = \psi_n(f) \circledast \psi_n(g)$. We interpret $\psi_n(f)$ as a partitioned matrix with $(n + 1) \times (n + 1)$ blocks, and each block is matrix of dimension of coefficient of f. With this view, $\psi_n(f) \circledast \psi_n(g)$ is matrix multiplication at the block level with the Hadamard multiplication of the entries. For matrix $\mathbf{c}$, $\psi_n(\mathbf{c} \cdot f) := \mathbf{c} \odot \psi_n(f)$, where $\mathbf{c} \odot M$ denotes matrix product of $\mathbf{c}$ with each block of the partitioned matrix M.

Definition 2.4 (ℓ-concentration in commutative case)**.** *If the coefficients of monomials in $f(\mathbf{X}, \mathbf{Y}) \in H_w(K)[\mathbf{X}, \mathbf{Y}]$, having $\mathbf{Y}$-degree greater than or equal to ℓ, have K-dependency on coefficients cor-*

responding to monomials with $\mathbf{Y}$*-degree at most* $\ell - 1$*, then we say that* f is ℓ-concentrated*.*

3 Proof for product depth one

A polynomial expressible as $\sum\prod$ circuit has sparsity at most s and [AJMR19] shows that it is nonzero on $\mathcal{M}_{\log s}(\mathbb{F})$ when $\mathbb{F}$ is sufficiently large. For the product of sparse polynomials, like $\prod\sum$ and $\prod\sum\prod$, consider the following simplified statement of Amitsur's theorem as explained in Makam's Ph.D. thesis [Mak18].

Theorem 3. *[Ami55, Ami72, Ami74] Any nonzero noncommutative rational expression, when substituted with generic* $d \times d$ *matrices is either identically zero, or invertible depending upon the value of* d*.*

When $f \not\equiv 0$, its sparse factors are nonzero. When substituted with generic matrices of $O(\log s)$ dimension these factors map to invertible matrices and hence their product is nonzero. Thus, $\prod\sum$ and $\prod\sum\prod$ are nonzero on $\mathcal{M}_{O(\log s)}(\mathbb{F})$.

This leaves us with the case of $\sum\prod\sum$ circuits. Let k and d be the maximum among the fan-in of the $\sum$-gates and the $\prod$-gates of the circuit respectively. For $\sum$-gates having fan-in less than k introduce $\prod$-gates that output the zero polynomial as child nodes to increase the fan-in to k. Similarly for $\prod$-gates having fan-in less than d introduce $\sum$-gates as its child nodes computing the constant polynomial 1. Thus, w.l.o.g. each product-gate has fanin d, and the top sum-fanin is k. It has the same depth, and the size is at most polynomial in the original size.

3.1 Reduction to the bivariate case

Proposition 3.1. *For any polynomial* $f \in \mathbb{F}\langle z_1, \ldots, z_n\rangle$ *of degree* d *having a homogeneous depth-*Δ *circuit of size* s*, the corresponding bivariate polynomial* $f'(x, y) \in \mathbb{F}\langle x, y\rangle$*, given by* $z_i \mapsto x^i y^{n-i}$*, has a homogeneous circuit of depth at most* $(\Delta + 1)$*, size* $O(sn)$ *and degree at most* nd *and is nonzero if and only if* f *is nonzero. Further, if* f *is given by a blackbox access for evaluation on matrices, we can efficiently create from it a blackbox access for* f'*.*

Proof. Since each z_i is mapped to a degree n monomial, if f was a homogeneous circuit then, f' will be a homogeneous circuit. Each monomial of degree d in $\mathbf{z}$ is now a monomial of degree nd in $\{x, y\}$. An extra layer of multiplication gates that compute $x^i y^{n-i}$ for all $i \in [n]$ feeds into leaves of the circuit of f to give a circuit for f'. The depth increases by one if the bottom most layer consisted of addition gates, otherwise the multiplication layers can be collapsed to a single layer. Thus the size of the circuit for f' is $O(sn)$.

Noncommutativity allows us to view a monomial as a string. Since each z_i maps to an n length string in x and y and because the map is one-one, it gives an encoding of the monomials in $\mathbf{z}$. Thus coefficients are preserved by the map and f is nonzero if and only if f' is nonzero.

Given a blackbox access for f, f' can be computed for any x, y matrices by evaluating f over $x^i y^{n-i}$. $\square$

Let C be a homogeneous depth-3 $\sum\prod\sum$ noncommutative formula $C(x,y) := \sum_{i=1}^{k}\prod_{j=1}^{d} (u_{ij}x + v_{ij}y)$ where $u_{ij}, v_{ij} \in \mathbb{F}$. Then,

$$C(x,y) = \mathbf{c} \cdot (u_1 x + v_1 y) \star \ldots \star (u_d x + v_d y)$$

where $\mathbf{c} = (1,\ldots,1)$, $u_j = (u_{1j},\ldots,u_{kj})^T \in H_k(\mathbb{F})$ and $v_j = (v_{1j},\ldots,v_{kj})^T \in H_k(\mathbb{F})$, $\cdot$ is the matrix product and $\star$ is the Hadamard product. It can be assumed that $\forall j \in [d], i \in [k], u_{ij} \neq 0$ and $v_{ij} \neq 0$ since if both are zero the i-th coordinate can be removed from the model, and if one of them is zero, the argument presented will work for $\pi(C)$ where $\pi : x \mapsto x + ry, y \mapsto x - ry$ for a random $r \in \mathbb{F}$. Since π is invertible $C \equiv 0 \iff \pi(C) \equiv 0$.

The corresponding commutative polynomial. Let $C(x,y) \in \mathbb{F}\langle x,y\rangle$ be a homogeneous $\sum\prod\sum$ noncommutative formula $C(x,y) = \mathbf{c} \cdot \prod_{i=1}^{n}(u_i x + v_i y)$. Define a corresponding set-multilinear commutative formula as $C^{\text{com}}(X_1,\ldots,X_n,Y_1,\ldots,Y_n) := \mathbf{c} \cdot \prod_{i=1}^{n}(u_i X_i + v_i Y_i)$ where $\{\mathbf{X},\mathbf{Y}\}$ are commuting indeterminates. Let m be a noncommuting monomial in $\{x,y\}$. Define a corresponding commutative monomial in $\{\mathbf{X},\mathbf{Y}\}$ as $m^{\text{com}} := \prod_{i=1}^{\deg(m)} z_i$ where $z_i = Y_i$ if $m(i) = y$ and $z_i = X_i$ if $m(i) = x$. Note that $\deg_{\mathbf{Y}}(m^{\text{com}}) = \deg_y(m)$.

Proposition 3.2. $C(x,y) \equiv 0 \iff C^{\text{com}}(\mathbf{X},\mathbf{Y}) \equiv 0$.

Proof. For all monomials m in $\{x,y\}$,

$$\operatorname{coef}(m,C) = \mathbf{c} \cdot \prod_{\substack{i\in[d],\\ m(i)=x}} u_i \star \prod_{\substack{i\in[d],\\ m(i)=y}} v_i = \mathbf{c} \cdot \prod_{\substack{i\in[d],\\ X_i \mid m^{\text{com}}}} u_i \star \prod_{\substack{i\in[d],\\ Y_i \mid m^{\text{com}}}} v_i = \operatorname{coef}(m^{\text{com}}, C^{\text{com}}).$$

Hence if C is nonzero, there is a nonzero coefficient for some monomial. The coefficient of the corresponding monomial in C^{com} is also nonzero. A similar argument proves the other direction. $\square$

3.2 Commutative results

Proposition 3.3. *Let $C := \mathbf{c} \cdot \prod_{i=1}^{d}(u_i X_i + v_i Y_i)$. For any monomial m in $\{\mathbf{X},\mathbf{Y}\}$ and any $S = \{s_1,\ldots,s_r\} \subset [d], s_{i-1} < s_i$, of size at most n, $\operatorname{coef}(m, \varphi_{n,S}(C))$ is a polynomial in $\{\mathbf{a},\mathbf{b}\}$. Any $w \in \operatorname{Supp}(\operatorname{coef}(m,\varphi_{n,S}(C)))$ in $\{\mathbf{a},\mathbf{b}\}$ has $\{a_{j,j}, b_{j,j}\}$-degree equal to $s_j - s_{j-1} - 1$. Consequently, for $|S| = |T|$ and $S \neq T$ and any monomial $m \in \operatorname{Supp}(C)$, $\operatorname{coef}(m,\varphi_{n,S}(C))$ and $\operatorname{coef}(m,\varphi_{n,T}(C))$ have disjoint support.*

Proof. First we prove this claim for $F := \prod_{i=1}^{d}(u_i X_i + v_i Y_i)$. Let $s_0 = 0, s_r = d+1$.

$$\varphi_{n,S} \circ F = \prod_{i=s_{j-1}+1}^{s_j - 1} \Big((u_i a_{j,j} + v_i b_{j,j}) X_i + v_i b_{j,j} Y_i \Big) \star G$$

where G is free of $a_{j,j}, b_{j,j}$. Here all the coefficients of $\{\mathbf{X}, \mathbf{Y}\}$ will have $\{a_{j,j}, b_{j,j}\}$-degree equal to $s_j - s_{j-1} - 1$. Since $\mathbf{c}$ is a matrix of constants $\operatorname{coef}(m, \varphi_{n,S}(C)) = \mathbf{c} \cdot \operatorname{coef}(m, \varphi_{n,S}(F))$ and it does not change the $\{a_{j,j}, b_{j,j}\}$-degree. Thus the claim holds for C.

Respectively when $T = \{t_1, \ldots, t_r\}$, all the coefficients in $\varphi_{n,T}(C)$ will have $\{a_{j,j}, b_{j,j}\}$-degree equal to $t_j - t_{j-1} - 1$. If k is the first index at which $s_k \neq t_k$, then $s_k - s_{k-1} - 1 \neq t_k - t_{k-1} - 1$ and the support will be disjoint. $\square$

Proposition 3.4. *For $u_i, v_i \in \mathbb{F}^w$ and $K = \mathbb{F}(\mathbf{a}, \mathbf{b}, \alpha)$, let $F := \prod_{i=1}^{d}(u_i X_i + v_i Y_i)$. Then, $\mathsf{span}\{\operatorname{coef}(m, \varphi_n(F))\} = \mathsf{span}\{\operatorname{coef}(m,\ F)\}$, where the span is over K. Consequently, setting α to a constant from $\mathbb{F}$ does not change the subspace.*

Proof. $\mathsf{span}\{\operatorname{coef}(\varphi_n(u_i X_i + v_i Y_i))\} = \{\eta_i u_i + \rho_i v_i, \rho_i v_i\} = \mathsf{span}\{u_i, v_i\}$ as η_i, ρ_i are in K.

When f_1, f_2 are over disjoint sets of variables $\operatorname{coef}(m_1 m_2, f_1 \star f_2) = \operatorname{coef}(m_1, f_1) \star \operatorname{coef}(m_2, f_2)$ where $m_1 \in \operatorname{Supp}(f_1)$ and $m_2 \in \operatorname{Supp}(f_2)$. The same is true for $\varphi_n(f_1), \varphi_n(f_2)$ in place of f_1, f_2.

If U_1, U_2, W_1, W_2 are subspaces such that $U_1 = U_2$ and $W_1 = W_2$ then $U_1 \times W_1 = U_2 \times W_2$ where by product of subspaces we mean the subspace generated by Hadamard product of all possible pairs of vectors with one vector from each subspace.

Hence, $\mathsf{span}\{\operatorname{coef}(m, \varphi_n(F))\} = \mathsf{span} \prod\{u_i, v_i\} = \mathsf{span}\{\operatorname{coef}(m, F)\}$.

The second statement is now obvious as the space is shown to be independent of α. $\square$

Lemma 3.5. *For $u_i, v_i \in H_w(\mathbb{F}), \ell := \lceil \log w \rceil + 1$, let $F := \prod_{i=1}^{d}(u_i X_i + v_i Y_i)$ and $\varphi := \varphi_\ell$. Then, $F' := \varphi(F)$ is ℓ-concentrated. Therefore if $C = \mathbf{c} \cdot F$, then $\varphi(C)$ is ℓ-concentrated.*

Remark 3.6. The proof also shows that for any $n > \ell$, $\varphi_n(F)$ and hence $\varphi_n(C)$ are ℓ-concentrated.

Proof. Let $K = \mathbb{F}(\mathbf{a}, \mathbf{b}, \alpha)$. The dependencies and rank in this proof are over K. Any monomial in $\{\mathbf{X}, \mathbf{Y}\}$ in the support of F can be uniquely written as $m = m_1 \ldots m_d$ where $m_i \in \{X_i, Y_i\}$. We call this a variable decomposition of m.

We will show that coefficient of a monomial m with $\deg_{\mathbf{Y}}(m) = t (> \ell)$ is in the span of coefficients of monomials with $\mathbf{Y}$-degree less than t in $\varphi(F)$. Let the variable decomposition of m be $m = m_1 \ldots m_d$ and T be a set of indices such that $|T| = \ell$ and $\forall i \in T,\ m_i = Y_i$. Let $F_T := \prod_{i \in T}(u_i X_i + v_i Y_i)$.

Finding a dependency. Consider a set of monomials $M := \prod_{i \in T}\{Y_i, X_i\} \cdot \prod_{i \in [d] \setminus T} m_i$. Since $|M| = 2^\ell > w$, there must exist a dependency among the coefficients corresponding to M in $\varphi_T(F)$. If the dependency does not have the coefficient of m, choose w such that it has the maximum $\mathbf{Y}$-degree. Let the dependency be:

$$\operatorname{coef}(w, \varphi_T \circ F) = \sum_{i=1}^{r} \beta_i \operatorname{coef}(w^{(i)}, \varphi_T \circ F)$$

where $\beta_i \in K, w^{(i)} \in M$.

Let $G_T := \prod_{i\in[d]\setminus T} \mathrm{coef}(m_i, \varphi_T(u_iX_i + v_iY_i))$, $G := \prod_{i\in[d]\setminus T} \mathrm{coef}(m_i, \varphi(u_iX_i + v_iY_i))$. They are both units since we assumed no coordinate of any u_i, v_i is zero. Every term in the dependency has G_T as a factor by construction of M. If the variable decompositions are $w = w_1 \dots w_d$ and $w^{(i)} = w_1^{(i)} \dots w_d^{(i)}$ then the dependency can also be written as

$$\mathrm{coef}\left(\prod_{i\in T} w_i, \varphi_T \circ F_T\right) \star G_T \;=\; \sum_{i=1}^{r} \beta_i \,\mathrm{coef}\left(\prod_{j\in T} w_j^{(i)}, \varphi_T \circ F_T\right) \star G_T\,.$$

Deriving a dependency with coefficient of m. The definition of φ implies $\varphi_T(F_T) = \prod_{t_i\in T} \big((a_{i,i+1}u_{t_i} + b_{i,i+1}v_{t_i})X_{t_i} + b_{i,i+1}v_{t_i}Y_{t_i}\big) =: \prod_{t_i\in T} (u'_{t_i}X_{t_i} + v'_{t_i}Y_{t_i})$ and that G_T does not have any variable from $\{b_{i,i+1} \mid i \in [\ell]\}$.

Observe the following properties of partial derivative operator:

$$\frac{\partial u'_{t_i}}{\partial b_{i,i+1}} = \frac{1}{b_{i,i+1}} \cdot v'_{t_i}$$
$$\frac{\partial v'_{t_i}}{\partial b_{i,i+1}} = \frac{1}{b_{i,i+1}} \cdot v'_{t_i}$$

This helps convert the coefficient of X_j to the coefficient of Y_j in $\varphi_T(F)$. To convert the coefficient of w to a coefficient of m with this operator, we need $\{b_{i,i+1}\}_i$ *uniquely* at the indices $\mathcal{Y}(m)\setminus\mathcal{Y}(w) =: S$. By the choice of M and the fact that $w \in M, S \subseteq T$. So, $\varphi_T(F)$ fulfills the criteria.

Let $\mathbf{b^e} := \prod_{s_j\in S} b_{j,j+1}$, then

$$\frac{\partial\, \mathrm{coef}(w, \varphi_T \circ F)}{\partial \mathbf{b^e}} = \beta \cdot \mathrm{coef}(m, \varphi_T \circ F), \qquad \beta \in K \setminus 0$$

Since $\deg_{\mathbf{Y}}(w) \geq \deg_{\mathbf{Y}}(w^{(i)})$, there exists a $r \in T$ such that $w_r = Y_r$ and $w_r^{(i)} = X_r$ in their variable decomposition. This means $r \notin S$ and so if

$$\frac{\partial\, \mathrm{coef}(w^{(i)}, \varphi_T \circ F)}{\partial \mathbf{b^e}} = \mathrm{coef}(w, \varphi_T \circ F),$$

then $w \in M \setminus m$ since w has X_r and $\prod_{i\in[d]\setminus T} w_i = \prod_{i\in[d]\setminus T} m_i$. As coefficients in the dependency, β_i, can also have $\{b_{i,i+1} \mid i \in [\ell]\}$ variables,

$$\frac{\partial \beta_i\, \mathrm{coef}(w^{(i)}, \varphi_T \circ F)}{\partial \mathbf{b^e}} \in \mathsf{span}\{\mathrm{coef}(w, \varphi_T \circ F) \mid w \in M \setminus m\}.$$

Therefore, differentiating the dependency by $\mathbf{b^e}$ gives

$$\mathrm{coef}(m, \varphi_T \circ F) \in \mathsf{span}\{\mathrm{coef}(w, \varphi_T \circ F) \mid \deg_{\mathbf{Y}}(w) < t\}.$$

Since G_T does not share any variables with $\mathbf{b^e}$ it is retained as a factor in the dependency and so the above is the same as

$$\text{coef}\,\Big(\prod_{i\in T} m_i, \varphi_T \circ F_T\Big) \star G_T \;\in\; \mathsf{span}\{\text{coef}(w, \varphi_T \circ F_T) \mid \deg_{\mathbf{Y}}(w) < \ell; w \in \text{Supp}(F_T)\} \star G_T\,.$$

Lifting a dependency on the coefficients of $\varphi_T(F)$ to a dependency on the coefficients of $\varphi(F)$. Since G_T is unit and $\prod_{i\in T} m_i = \prod_{i\in T} Y_i$, we have proved that for any T of size ℓ,

$$\mathsf{span}\{\text{coef}(w, \varphi_T \circ F_T) \mid w \in \text{Supp}(\varphi_T \circ F_T)\} = \mathsf{span}\{\text{coef}(w, \varphi_T \circ F_T) \mid \deg_{\mathbf{Y}}(w) < \ell\}. \tag{3.7}$$

If we denote $\mathcal{W} := \mathsf{span}\{\text{coef}(w, \varphi \circ F_T) \mid w \in \text{Supp}(\varphi \circ F_T)\}$ and $\mathcal{U} := \mathsf{span}\{\text{coef}(w, \varphi \circ F_T) \mid \deg_{\mathbf{Y}}(w) < \ell\}$ then Equation 3.7 is can be restated as $\mathcal{W}|_{\alpha=\alpha_T} = \mathcal{U}|_{\alpha=\alpha_T}$.

Observe, $\text{rk}\,\mathcal{U}|_{\alpha=\alpha_T} \leq \text{rk}\,\mathcal{U} \leq \text{rk}\,\mathcal{W} = \text{rk}\,\mathcal{W}|_{\alpha=\alpha_T}$. The first inequality is true because the rank cannot increase by α substitution. Consider a matrix with columns as vectors of $\mathcal{U}|_{\alpha=\alpha_T}$. A nonzero minor here implies a corresponding nonzero minor in a matrix with columns as vectors of $\mathcal{U}$. The second inequality is justified as $\mathcal{U}$ is a subspace of $\mathcal{W}$ and Proposition 3.4 proves the third. Thus $\text{rk}\,\mathcal{U} = \text{rk}\,\mathcal{W}$ and $\mathcal{U} = \mathcal{W}$.

Therefore, now without the α substitution we have,

$$\mathsf{span}\{\text{coef}(w, \varphi \circ F_T) \mid w \in \text{Supp}(\varphi \circ F_T)\} = \mathsf{span}\{\text{coef}(w, \varphi \circ F_T) \mid \deg_{\mathbf{Y}}(w) < \ell; w \in \text{Supp}(\varphi \circ F_T)\}.$$

Thus $\text{coef}(\prod_{i\in T} m_i, \varphi \circ F_T) \star G_T \in \mathsf{span}\{\text{coef}(w, \varphi \circ F_T) \mid \deg_{\mathbf{Y}}(w) < \ell; w \in \text{Supp}(\varphi \circ F_T)\} \star G_T$. Hadamard product with $G_T^{-1} \star G$, gives $\text{coef}(m, \varphi \circ F) \in \mathsf{span}\{\text{coef}(w, \varphi \circ F) \mid \deg_{\mathbf{Y}}(w) < t; w \in \text{Supp}(\varphi \circ F)\}$.

Finally, induction on t gives ℓ concentration for $\varphi \circ F$.

Any dependency showing $\text{coef}(m, \varphi \circ F) \in \mathsf{span}\{\text{coef}(w, \varphi \circ F) \mid \deg_{\mathbf{Y}}(w) < \ell\}$, can be multiplied by $\mathbf{c}$ to show $\text{coef}(m, \varphi \circ C) \in \mathsf{span}\{\text{coef}(w, \varphi \circ C) \mid \deg_{\mathbf{Y}}(w) < \ell\}$. Thus $\varphi(C)$ is also ℓ-concentrated. □

Lemma 3.8. *If $\varphi_\ell(C)$ is ℓ-concentrated then,*
$\forall m, \deg_{\mathbf{Y}}(m) < \ell, \forall S \in \binom{[d]}{\ell}, \text{coef}(m, \varphi_{\ell,S}(C)) = 0 \implies \varphi_\ell(C) \equiv 0.$

Proof. Consider the statements,

$$\begin{aligned} \varphi_\ell(C) \equiv 0 &\iff \forall m, \deg_y(m) < \ell, \text{coef}(m, \varphi_\ell(C)) = 0 \\ &\iff \forall S \subseteq [d], |S| \leq \ell, \text{coef}(m, \varphi_{\ell,S}(C)) = 0 \end{aligned}$$

The forward direction in the first if and only if statement is trivial and the backward implication was proved in Lemma 3.5. The second statement is true by the property of interpolating polynomials used in the definition of φ_ℓ. To see that all sets of size exactly ℓ are sufficient, consider $T \subsetneq S$. If $\text{coef}(m, \varphi_{\ell,S}(C)) = 0$ then $\text{coef}(m, \varphi_{\ell,T}(C)) = 0$ because there exists a mapping of $\{\mathbf{a}, \mathbf{b}\}$ that converts the former to the latter. It is sufficient to see this map for $T = \{s_1, \ldots, s_{k-1}, s_k, \ldots, s_r\} \subsetneq \{s_1, \ldots, s_r\} = S$. The map is,

$a_{k,k+1} \mapsto a_{k,k}$, $b_{k,k+1} \mapsto b_{k,k}$,

$\forall i, j > k, a_{i,j} \mapsto a_{i-1,j-1}$ and $b_{i,j} \mapsto b_{i-1,j-1}$,

and identity on the rest.

This proves the claim. $\square$

3.3 Noncommutative results

Matrix substitutions of x and y. We denote a $(n+1)\times(n+1)$ matrix with commuting variables in diagonal and super-diagonal entries and zero elsewhere by

$$M_{\mathbf{x},n} = \begin{bmatrix} x_{1,1} & x_{1,2} & 0 & \ldots & 0 & 0 \\ 0 & x_{2,2} & x_{2,3} & \ldots & 0 & 0 \\ 0 & 0 & x_{3,3} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & x_{n,n} & x_{n,n+1} \\ 0 & 0 & 0 & \ldots & 0 & x_{n+1,n+1} \\ . \end{bmatrix}$$

A similar matrix with $y_{i,j}$ instead of $x_{i,j}$ is denoted by $M_{\mathbf{y},n}$. Consider a noncommutative monomial m and the map $x \mapsto M_{\mathbf{x},n}, y \mapsto M_{\mathbf{y},n}$. Let $m(M_{\mathbf{x},n}, M_{\mathbf{y},n})$ denote the result of application of this map. By the definition of matrix multiplication and the choice of $M_{\mathbf{x},n}, M_{\mathbf{y},n}$, it can be observed that the $(1, \ell+1)$-th entry of $m(M_{\mathbf{x},n}, M_{\mathbf{y},n})$ is the sum of monomials produced by different paths from q_1 to q_ℓ in the substitution NFA graph in Figure 1. Each step on a path consumes a character of m i.e. x or y and accordingly outputs a variable from $\{x_{i,j}, y_{i,j} \mid i, j \in [n+1]\}$. The variables along a path are multiplied to produce a monomial for this path. We refer the reader to [AJMR19] for a thorough explanation. It is then also easy to see that the $(1, t+1)$-th entry of

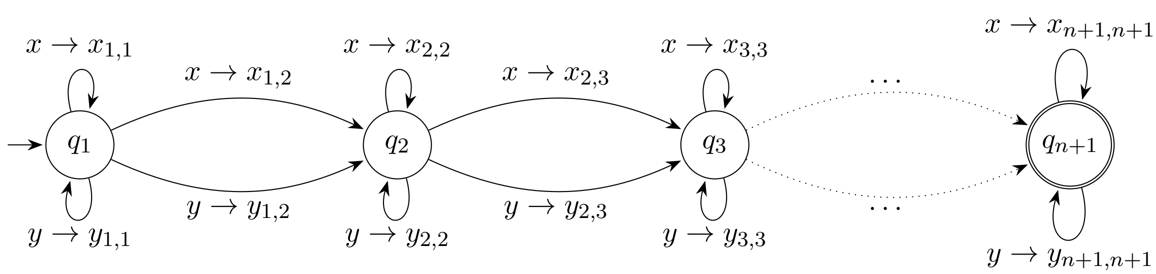

Figure 1: Transition diagram of the substitution NFA.

$m(M_{\mathbf{x},n}, M_{\mathbf{y},n})$ is the sum of all possible outputs produced by paths from q_1 to q_t. The $(1, t+1)$-th entry of $f(M_{\mathbf{x},n}, M_{\mathbf{y},n})$ is then the sum of all $(1, t+1)$-th entries of $m(M_{\mathbf{x},n}, M_{\mathbf{y},n})$ accompanied by the coefficient of m in f.

We call the transition steps between states as jump and the rest as loops. For a monomial, a path is uniquely identified when we specify the index of the jump steps. On jumps the output variables are $\{x_{i,i+1}, y_{i,i+1} \mid i \in [n]\}$ and on loops the output variables are $\{x_{i,i}, y_{i,i} \mid i \in [n+1]\}$. For a monomial with y-degree less than ℓ, consider a path that jumps on seeing a y. The commutative monomial produced by this path is unique i.e. it cannot be produced by any other monomial. This allows us to isolate the coefficients of such monomials.

Lemma 3.9 (Matrix substitutions of x and y)**.** *If there exists a monomial m in the support of $f \in \mathbb{F}\langle x, y\rangle$ with $\deg_y(m) < \ell$, then $f(M_{\mathbf{x},\ell}, M_{\mathbf{y},\ell})$ is a nonzero matrix.*

Proof. If for every monomial m, with $\deg_y(m) < \ell$, the corresponding matrix product $\prod_{i=1}^{\deg m} M_{m(i),\ell}$ $=: m(M_{\mathbf{x},\ell}, M_{\mathbf{y},\ell})$ produces a unique commutative monomial in one of its entries, then it preserves the coefficient of m in f.

For $m := x^{e_1}yx^{e_2}y\dots x^{e_t}yx^{e_{t+1}}$, $t \le \ell$, the commutative monomial $x_{1,1}^{e_1}y_{1,2}x_{2,2}^{e_2}y_{2,3}\dots x_{t,t}^{e_t}y_{t,t+1}$ $x_{t+1,t+1}^{e_{t+1}}$ can only occur in $m(M_{\mathbf{x},\ell}, M_{\mathbf{y},\ell})$. Thus some entry of $f(M_{\mathbf{x},\ell}, M_{\mathbf{y},\ell})$ is guaranteed to be nonzero if f has a nonzero coefficient for some monomial having y-degree less than ℓ. $\square$

Properties of the transformation ψ. There is an interesting relation between the entries of coefficient matrix of $\psi_n(C)$ and its commutative counterpart $\varphi_n(C^{\text{com}})$.

Proposition 3.10. *Let $F = \prod_{i=1}^{d}(u_i x + v_i y)$ with $u_i, v_i \in H_w(\mathbb{F})$ and $C = \mathbf{c} \cdot F$.*

1. *The $(1, t+1)$-th entry of the block matrix $\operatorname{coef}(m, \psi_n \circ F)$ is*

   $\operatorname{coef}(m, \psi_n \circ F)_{1,t+1} = \sum_S \operatorname{coef}(m^{\text{com}}, \varphi_{n,S}(F^{\text{com}}))$ *where summation is over all possible $S \in \binom{[d]}{t}$.*

2. *The (p, q)-th entry of the block matrix $\operatorname{coef}(m, \psi_n \circ F)$ is*

   $\operatorname{coef}(m, \psi_n \circ F)_{p,q} = \sigma_{p-1} \circ \operatorname{coef}(m, \psi_n \circ F)_{1,q-p+1}$ *where $\sigma_{p-1} : a_{i,j} \mapsto a_{i+p-1,j+p-1}$ and $b_{i,j} \mapsto b_{i+p-1,j+p-1}$.*

3. *The $(1, t+1)$-th entry of the block matrix $\operatorname{coef}(m, \psi_n \circ C)$ is*

   $\operatorname{coef}(m, \psi_n \circ C)_{1,t+1} = \sum_S \operatorname{coef}(m^{\text{com}}, \varphi_{n,S}(C^{\text{com}}))$ *where summation is over all possible $S \in \binom{[d]}{t}$.*

Proof. Let $P := \operatorname{coef}(m, \psi_n(F)) = Z^{(1)} \circledast \dots \circledast Z^{(d)}$ where $Z^{(i)} = u_i \otimes A + v_i \otimes B$ if $m(i) = x$ and $Z^{(i)} = v_i \otimes B$ if $m(i) = y$. Interpret $Z^{(k)}$ as a block matrix with entries $z_{i,j}^{(k)} \in H_w(\mathbb{F})[\mathbf{a}, \mathbf{b}]$. From the definition of $\circledast$, matrix multiplication for outer partitioned matrix gives

$$P_{1,t+1} = \sum_{\substack{\{r_1,\dots,r_d\} \in [n]^d \\ r_1 = 1, r_d = t+1}} \prod_{i=1}^{d} z_{r_i,r_{i+1}}^{(i)}, \text{where, the product on } z_{i,j}^{(k)} \text{ is the Hadamard product.}$$

$z_{i,j}^{(k)}$ is nonzero only for $z_{i,i}^{(k)}$ and $z_{i,i+1}^{(k)}$ by definition of A, B. So the summation is reduced to be over $\{r_1, \ldots, r_d\} \in [n]^d$ such that $r_1 = 1, r_d = t+1$ and $\forall 1 < i < d$, $r_{i+1} = r_i$ or $r_{i+1} = r_i + 1$.

Fix an $R := \{r_1, \ldots, r_d\}$ as above and define $S := \{i \mid r_{i+1} = r_i + 1\}$. It is easy to reason $|S| = t$. If $s_1, \ldots, s_t$ represent the elements of S arranged in ascending order, with $s_0 = 0, s_{t+1} = d$

$$\prod_{i=1}^{d} z_{r_i,r_{i+1}}^{(i)} = \prod_{j=1}^{t} z_{j,j+1}^{(s_j)} \star \prod_{\substack{i \notin S \\ s_{k-1} < i < s_k}} z_{k,k}^{(i)} = \mathrm{coef}(m, \varphi_{n,S}(F^{\mathrm{com}})).$$

The last equality follows from the definition of $\mathrm{coef}(m, \varphi_{n,S}(F^{\mathrm{com}}))$. Summation over all valid $\{r_1, \ldots, r_d\}$ translates to summation over all possible subsets S of $[d]$ of size t. This proves the first statement.

For the second statement consider $(1, q-p+1)$-th entry of P. Here the summation is over $R = \{r_1, \ldots, r_d\} \in [n]^d$ such that $r_1 = 1, r_d = q-p+1$ and $\forall 1 < i < d$, $r_{i+1} = r_i$ or $r_{i+1} = r_i + 1$. There is a corresponding $R' = \{r_1 - p + 1, \ldots, r_d - p + 1\}$ which contributes to (p, q)-th entry of P. The proof of the second statement now follows from the definition of $z_{r_i,r_{i+1}}^{(i)}$ which is either $u_i a_{r_i,r_{i+1}} + v_i b_{r_i,r_{i+1}}$ or $v_i b_{r_i,r_{i+1}}$.

For the third statement, from the definition of ψ_n, $\mathrm{coef}(m, \psi_n(C))_{1,t+1} := \mathbf{c} \cdot \mathrm{coef}(m, \psi_n(F))_{1,t+1}$. Using the first statement, $\mathbf{c} \cdot \mathrm{coef}(m, \psi_n(F))_{1,t+1} = \mathbf{c} \cdot \sum_S \mathrm{coef}(m^{\mathrm{com}}, \varphi_{n,S}(F^{\mathrm{com}})) = \sum_S \mathrm{coef}(m^{\mathrm{com}}, \varphi_{n,S}(\mathbf{c} \cdot F^{\mathrm{com}})) = \sum_S \mathrm{coef}(m^{\mathrm{com}}, \varphi_{n,S}(C^{\mathrm{com}}))$. Since $C^{\mathrm{com}} = \mathbf{c} \cdot F^{\mathrm{com}}$ and $\mathbf{c}$ is matrix of constants, these statements are true. $\square$

Proposition 3.10 can also be understood with substitution NFA. First consider a product of d matrices each of which is A or $A + B$. An entry of the product matrix is a sum over paths that produce $a_{i,i+1} + b_{i,i+1}$ or $b_{i,i+1}$ on jumps and $a_{i,i} + b_{i,i}$ or $b_{i,i}$ on loops in a graph like in Figure 1.

Now consider the product $\prod Z_i$ where $Z_i = u_i \otimes A + v_i \otimes B$ if $m(i) = x$ and $Z_i = v_i \otimes B$ if $m(i) = y$. This is $\mathrm{coef}(m, \psi_n(F))$. Since u_i and v_i correspond to specific index positions in m, the commutative variant m^{com} that remembers the indices of m will be useful. The $(1, t+1)$-th entry of $\mathrm{coef}(m, \psi_n(F))$ is the sum of outputs produced by the substitution NFA in Figure 2 on all paths from state q_1 to state q_{t+1} that consume a character of the string m^{com} at each step. If we fix a path with jumps at indices in the set S, the output on this path is $\mathrm{coef}(m^{\mathrm{com}}, \varphi_S(F^{\mathrm{com}}))$.

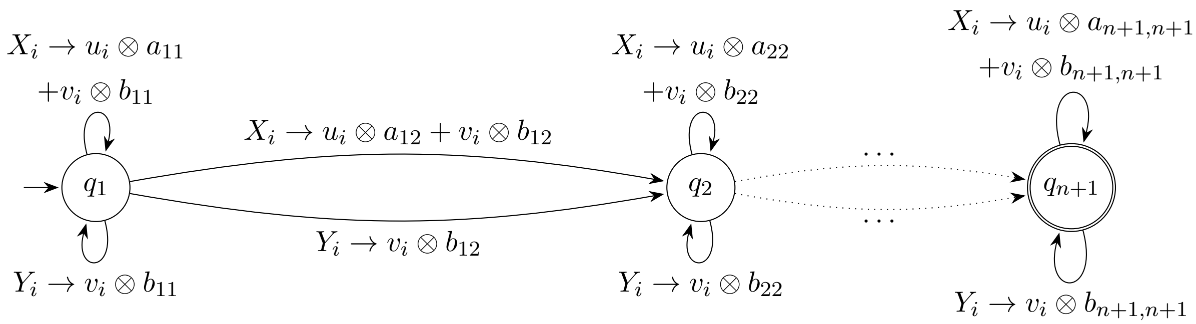

Figure 2: Transition diagram of the substitution NFA.

3.4 Concentration in noncommutative case

Lemma 3.11. *For $u_i, v_i \in H_w(\mathbb{F}), \ell := \lceil \log w \rceil + 1$, let $C := \mathbf{c} \cdot \prod_{i=1}^{d} (u_i x + v_i y)$. Then $C \equiv 0 \iff \forall m, \deg_y(m) < \ell, \operatorname{coef}(m, \psi_\ell(C)) = \mathbf{0}$.*

Remark 3.12. For the proof it is crucial to have $\mathbf{c} \in \mathbb{F}^w$ (i.e., each coordinate is a constant, in $\mathbb{F}$, free of ψ_ℓ variables) otherwise the path argument using S breaks down.

Proof. Only the backward direction is non-trivial. Assume $\forall m, \deg_y(m) < \ell, \operatorname{coef}(m, \psi_\ell(C)) = \mathbf{0}$. In particular, for all such monomials, the $(1, \ell+1)$-th entry is a zero.

The third statement of Proposition 3.10 shows,

$\operatorname{coef}(m, \psi_\ell(C))_{1,\ell+1} = \sum_{S \in \binom{[d]}{\ell}} \operatorname{coef}(m^{\text{com}}, \varphi_{\ell,S}(C^{\text{com}}))$.

Since for different sets S, $\operatorname{coef}(m^{\text{com}}, \varphi_{\ell,S}(C^{\text{com}}))$ have disjoint support (Proposition 3.3),coef($m^{\text{com}}, \varphi_{\ell,S}(C^{\text{com}})) = 0$ for all subsets S.

Thus we have $\forall m^{\text{com}}, \deg_{\mathbf{Y}}(m^{\text{com}}) < \ell, \forall S \in \binom{[d]}{\ell}, \operatorname{coef}(m^{\text{com}}, \varphi_{\ell,S}(C^{\text{com}})) = 0$. This together with Lemma 3.5 which proves ℓ-concentration of $\varphi_\ell(C^{\text{com}})$ implies $\varphi_\ell(C^{\text{com}}) \equiv 0$ (Lemma 3.8). As φ_ℓ is an invertible transformation, $C^{\text{com}} \equiv 0$. We finally arrive at the conclusion in the lemma using $C^{\text{com}} \equiv 0 \iff C \equiv 0$ (Proposition 3.2). $\square$

Proof of Theorem 1 for bivariate polynomial computed by a $\sum\prod\sum$ circuit. As stated at the beginning of this section a noncommutative degree d homogeneous $\sum\prod\sum$ circuit can be expressed as $C(x, y) := \sum_{i=1}^{k} \prod_{j=1}^{d} \ell_{i,j}(x, y) = \mathbf{c}^T \cdot f_1 \star \ldots \star f_d$ where $f_i \in H_k(\mathbb{F})\langle x, y\rangle$ are linear forms and $\mathbf{c} \in \mathbb{F}^k$. If the size of the circuit is s, the degree d is at most $\operatorname{poly}(s)$.

Lemma 3.11 shows that for $\ell = O(\log k)$, it is enough to check the nonzeroness of the coefficients corresponding to monomials with y-degree less than ℓ in $\psi_\ell(C(x, y))$. This can be done by substituting x and y with matrices $M_{\mathbf{x},\ell}$ and $M_{\mathbf{y},\ell}$ as shown in Lemma 3.9.

Therefore, if $C(x, y) \not\equiv 0$ then $C(A \otimes M_{\mathbf{x},\ell}, B \otimes (M_{\mathbf{x},\ell} + M_{\mathbf{y},\ell}) \not\equiv 0$. Here A, B as well as $M_{\mathbf{x},\ell}, M_{\mathbf{y},\ell}$ are $(\ell+1) \times (\ell+1)$ matrices. The final map is a substitution of x and y by symbolic matrices of dimension $O(\log^2 k) = O(\log^2 s) (\because k \leq s)$. The output is a matrix with entries that are polynomials of degree $2d$ in the commutative variables $\{a_{ij}, b_{ij}, x_{ij}, y_{ij} \mid i, j \in [\ell+1]\}$. The polynomial identity lemma Lemma 2.1 over an extension field of dimension $O(\log s)$ for a nonzero entry of the matrix gives the PIT algorithm. In conclusion, the blackbox randomized PIT algorithm is to substitute x, y with matrices of dimension $O(\log^2 s)$ with random entries from an extension field of degree $O(\log s)$. $\square$

4 Proof for product depth two

Let C be a polynomial computed by $\sum\prod\sum\prod\sum$ circuit over $\mathbb{F} = H_{k^0}(\mathbb{F})$ with product gate fan-in d and sum gate fan-in k.

$$
\begin{aligned}
C &= \sum_{i=1}^{k}\prod_{j=1}^{d} g_{ij} \\
&= \mathbf{c}_1 \cdot f_1 \star f_2 \star \ldots \star f_d
\end{aligned}
$$

where $f_j = (g_{1j}, g_{2j}, \ldots, g_{kj})^T \in \mathbb{F}^k\langle x, y\rangle$, $\mathbf{c}_1 = \mathbf{1}_k^T$ and each g_{ij} is a $\sum\prod\sum$ circuit. Thus f_j is a $\sum\prod\sum$ circuit over $H_k(\mathbb{F})$.

Any f_j is structurally like f:

$$
\begin{aligned}
f &= \sum_{i=1}^{k}\prod_{j=1}^{d}(u_{ij}x + v_{ij}y), \quad \text{with } u_{ij}, v_{ij} \in \mathbb{F}^k \\
&= \mathbf{c} \cdot \prod_{j=1}^{d} \begin{bmatrix} u_{1j}x + v_{1j}y \\ u_{2j}x + v_{2j}y \\ \vdots \\ u_{kj}x + v_{kj}y \end{bmatrix}_{k^2\times 1} \\
&= \mathbf{c} \cdot \prod_{i=1}^{d}(U_i x + V_i y)
\end{aligned}
$$

where $\mathbf{c} = \begin{bmatrix} e_1 & e_1 & \ldots & e_1 \\ e_2 & & \ldots & e_2 \\ \vdots & & & \\ e_k & & \ldots & e_k \end{bmatrix}_{k\times k^2}$, $\{e_i\}_i$ is the standard basis of the row-space $\mathbb{R}^k$, $U_j = (u_{1j}^T, \ldots, u_{kj}^T)^T$, $V_j = (v_{1j}^T, \ldots, v_{kj}^T)^T$, $U_j, V_j \in H_{k^2}(\mathbb{F})$. If there are constants on the wires, the matrix $\mathbf{c}$ can be modified accordingly.

In summary, a depth-2 circuit has the form $C = \mathbf{c}_1 \cdot f_1 \star \ldots \star f_d$, where each f_i has the form $f_i = \mathbf{c} \cdot \prod_{j=1}^{d}(U_{ij} \cdot x + V_{ij} \cdot y)$ for some $U_{ij}, V_{ij} \in H_{k^2}(\mathbb{F})$.

The corresponding commutative polynomial. Like in the product depth-1 case we now define a corresponding commutative polynomial. The base case of the inductive definition is like in Proposition 3.2: if $D = \prod_{i=1}^{n}(u_i x + v_i y)$ then $D^{\text{com}} := \prod_{i=1}^{n}(u_i X_i + v_i Y_i)$. For $f = \mathbf{c} \cdot g$ we define $f^{\text{com}} = \mathbf{c} \cdot g^{\text{com}}$ and for $f = g \star h$ we define $f^{\text{com}} = g^{\text{com}} \star \tau_d(h^{\text{com}})$, where $d := \deg g$ and τ_d maps $X_i \mapsto X_{i+d}$ and $Y_i \mapsto Y_{i+d}$. This ensures that g^{com} and h^{com} are on disjoint sets of variables. The noncommutative monomial m maps to m^{com} under this map. This will define C^{com} for the C above. This also defines a corresponding commutative polynomial for a homogeneous circuit of any depth. Collect all such commutative polynomials in the class $\mathcal{C}$.

Definition 4.1. *The class of commutative polynomials $\mathcal{C}$ is defined inductively as follows. In the base case, $\mathcal{C}$ contains all linear forms $(uX_1 + vY_1)$, where u and v are vectors over the field $\mathbb{F}$. The class is closed under a variable disjoint Hadamard product: for any $f, g \in \mathcal{C}$ with $d = \deg(f)$, the polynomial $f \star \tau_d(g)$ belongs to $\mathcal{C}$, where τ_d is the map $X_i \mapsto X_{i+d}$ and $Y_i \mapsto Y_{i+d}$. Finally, if $f \in \mathcal{C}$, then $\mathbf{c} \cdot f \in \mathcal{C}$ for any matrix $\mathbf{c}$ of suitable dimensions over $\mathbb{F}$.*

Any monomial in $\{\mathbf{X}, \mathbf{Y}\}$ in the support of $C(\in \mathcal{C})$ can be uniquely decomposed as $m = \hat{m}_1 \dots \hat{m}_d$ where $\hat{m}_i \in \text{Supp}(f_i)$, because f_i's are over disjoint sets of variables. We call such a decomposition as *block decomposition of m*. We denote by $\text{bdeg}_{\mathbf{Y}}(m)$ the number of $\hat{m}_i$ with $\deg_{\mathbf{Y}}(\hat{m}_i) \geq 1$.

We prove a result analogous to Proposition 3.2 but now for a product depth-2 circuit.

Proposition 4.2. $C(x, y) \equiv 0 \iff C^{\text{com}}(\mathbf{X}, \mathbf{Y}) \equiv 0$.

Proof. Let $C = \mathbf{c} \cdot f_1 \star \dots \star f_d$. We know from Proposition 3.2, $\forall i \in [d], \forall m \in \text{Supp}(f_i)\, \text{coef}(m, f_i) = \text{coef}(m^{\text{com}}, f_i^{\text{com}})$. Let for the following argument $d = \deg(h), \tau_d : X_i \mapsto X_{i+d}, Y_i \mapsto Y_{i+d}$.

$$\begin{aligned}
\text{coef}(m_1 m_2, h \star g) &= \text{coef}(m_1, h) \star \text{coef}(m_2, g) && \text{by noncommutativity.} \\
&= \text{coef}(m_1^{\text{com}}, h^{\text{com}}) \star \text{coef}(m_2^{\text{com}}, g^{\text{com}}) && \text{by induction on } h, g. \\
&= \text{coef}(m_1^{\text{com}}, h^{\text{com}}) \star \text{coef}(\tau_d \circ m_2^{\text{com}}, \tau_d \circ g^{\text{com}}) \\
&= \text{coef}(m_1^{\text{com}}(\tau_d \circ m_2^{\text{com}}), h^{\text{com}} \star \tau_d \circ g^{\text{com}}) && h, \tau_d \circ g \text{ are on disjoint variables} \\
&= \text{coef}((m_1 m_2)^{\text{com}}, (h \star g)^{\text{com}}) && \text{definition of } C^{\text{com}}
\end{aligned}$$

$\text{coef}(m_2^{\text{com}}, g^{\text{com}}) = \text{coef}(\tau_d \circ m_2^{\text{com}}, \tau_d \circ g^{\text{com}})$ since renaming all variables in g^{com} and correspondingly in m^{com} does not change the monomial.

With this argument we have $\text{coef}(m, f_1 \star \dots \star f_d) = \text{coef}(m^{\text{com}}, (f_1 \star \dots \star f_d)^{\text{com}})$. Matrix multiplication with $\mathbf{c}$ on both sides, proves the claim for C. $\square$

4.1 Commutative results

The results in this section are for the commutative polynomial derived from a noncommutative homogeneous product depth-2 circuit in $\{x, y\}$ as defined above. All the proofs in this subsection are for the the polynomial that can be expressed as $C(\mathbf{X}, \mathbf{Y}) = \mathbf{c} \cdot f_1 \star \dots \star f_d$, where $f_i = \mathbf{c}_i \cdot \prod_{j=n(i-1)+1}^{ni} u_j X_j + v_j Y_j$, $u_j, v_j \in H_{w_1}(\mathbb{F}), \mathbf{c}_i \in \mathbb{F}^{w_2 \times w_1}$ and $\mathbf{c} \in \mathbb{F}^{w_3 \times w_2}$. For a monomial m in the support of f_i, $\text{coef}(m, f_i)$ is a vector of dimension w_2.

Let K be the functional field $\mathbb{F}(\mathbf{a}, \mathbf{b}, \alpha)$ where $\mathbf{a}, \mathbf{b}, \alpha$ are the variables in the definition of φ_n.

Proposition 4.3 and *Proposition* 4.4 we prove next are analogous to Proposition 3.3 and *Proposition* 3.4 respectively, for the product depth-2.

Proposition 4.3. *Let $C := \mathbf{c} \cdot f_1 \star \ldots \star f_d$. For any monomial m in $\{\mathbf{X}, \mathbf{Y}\}$ and any $S = \{s_1, \ldots, s_r\} \subset [d], s_{i-1} < s_i$, of size at most n, $\text{coef}(m, \varphi_{n,S}(C))$ is a polynomial in $\{\mathbf{a}, \mathbf{b}\}$. Any $w \in \text{Supp}(\text{coef}(m, \varphi_{n,S}(C)))$ in $\{\mathbf{a}, \mathbf{b}\}$ has $\{a_{j,j}, b_{j,j}\}$-degree equal to $s_j - s_{j-1} - 1$. Consequently, for $|S| = |T|$ and $S \neq T$ and any monomial $m \in \text{Supp}(C)$, $\text{coef}(m, \varphi_{n,S}(C))$ and $\text{coef}(m, \varphi_{n,T}(C))$ have disjoint support.*

Proof. The base case, $\text{coef}(m, \varphi_{n,S}(f_i))$ was proved in Proposition 3.3.

Let $f = g \star h$, where g, h are on disjoint variables in $\{\mathbf{X}, \mathbf{Y}\}$ and $d_g = \deg(g)$. If $s_j \leq d_g$ then $\varphi_{n,S}(h)$ will not have any $\{a_{j,j}, b_{j,j}\}$ variables and the claim holds by induction on $\varphi_{n,S}(g)$. Similarly if $d_g \leq s_{j-1}$ the claim holds by induction on $\varphi_{n,S}(h)$.

If $s_{j-1} < d_g < s_j$, then all monomials in the support of $\varphi_{n,S}(g)$ have $\{a_{j,j}, b_{j,j}\}$-degree equal to $d_g - s_{j-1}$ and all monomials in the support of $\varphi_{n,S}(h)$ have $\{a_{j,j}, b_{j,j}\}$-degree equal to $d_g - s_j - 1$. This proves the claim for $f_1 \star \ldots \star f_d$.

Matrix product with a constant matrix does not change the monomials. Hence the claim holds for C.

Respectively when $T = \{t_1, \ldots, t_r\}$, all the coefficients in $\varphi_{n,T}(C)$ will have $\{a_{j,j}, b_{j,j}\}$-degree equal to $t_j - t_{j-1} - 1$. If k is the first index at which $s_k \neq t_k$, then $s_k - s_{k-1} - 1 \neq t_k - t_{k-1} - 1$ and the support will be disjoint. $\square$

Proposition 4.4. *For any $T \subseteq [d]$, and $D_T := \prod_{i \in T} f_i$, $\mathsf{span}\{\text{coef}(m, \varphi_n \circ \prod_{i \in T} f_i), \mid m \in \text{Supp}(D_T)\} = \mathsf{span}\{\text{coef}(m, D_T) \mid m \in \text{Supp}(D_T)\}$ where the span is over K. Consequently, setting α to value from $\mathbb{F}$ does not change the subspace.*

Proof. The proof is based on the recursive definition of the commutative polynomial. For the base case like $f_i = \mathbf{c} \cdot \prod_{j=n(i-1)+1}^{ni} (u_j X_j + v_j Y_j)$ Proposition 3.4 proved $\mathsf{span}\{\text{coef}(m, \varphi_n \circ f_i)\} = \mathsf{span}\{\text{coef}(m, f_i)\}$ for all $i \in [T]$.

For monomial $m \in \text{Supp}(D_T)$ with block decomposition $m = \prod_{i \in T} \hat{m}_i$, $\hat{m}_i \in \text{Supp}(f_i)$, $\text{coef}(m, \prod_{i \in T} f_i) = \prod_{i \in T} \text{coef}(\hat{m}_i, f_i)$ because f_i's are on disjoint variables $\{\mathbf{X}, \mathbf{Y}\}$. Since φ_n is a homomorphism and $\varphi_n(f_i)$'s are on disjoint variables, $\text{coef}(m, \varphi_n \circ \prod_{i \in T} f_i) = \prod_{i \in T} \text{coef}(m_i, \varphi_n(f_i))$.

If U_1, U_2, W_1, W_2 are subspaces such that $U_1 = U_2$ and $W_1 = W_2$ then $U_1 \times W_1 = U_2 \times W_2$ where by product of subspaces we mean the subspace generated by Hadamard product of all possible pairs of vectors with one vector from each subspace. Therefore, $\mathsf{span}\{(\text{coef}(m, \prod_{i \in T} \varphi_n(f_i))\} = \mathsf{span}\{\text{coef}(m, D_T)\}$.

This proves the first claim and the second claim follows, as now the subspace is shown to be independent of α. $\square$

The concentration result for product depth-1 in Lemma 3.5 used a partial derivative operator on dependencies. The next result shows that this operator gives us the desired properties also on coefficients from product depth-2 circuits.

Lemma 4.5. *Let $F = \prod_{i\in[k]} f_i$, be a commutative polynomial of degree d and $\varphi = \varphi_n$ for some n. Let $S := \{s_1, \ldots, s_t\}, t < n$, be some subset of $[d]$, and $P \subseteq S$. Define $\mathbf{b}^{\mathbf{e}} := \prod_{s_j \in S\cap P} b_{j,j+1}$. Then, for the monomial w in $\{\mathbf{X}, \mathbf{Y}\}$ with variable decomposition $w = w_1 \ldots w_d$,*

$$\frac{\partial\, \mathrm{coef}(w, \varphi_S \circ F)}{\partial \mathbf{b}^{\mathbf{e}}} = \beta \cdot \mathrm{coef}(m, \varphi_S \circ F), \;\; \beta \in K \setminus 0$$

where $m = \prod_{i=1}^{[d]} Z_i$, $Z_i = w_i$ if $i \notin S \cap P$ and $Z_i = Y_i$ if $i \in S \cap P$.

Proof. Since F is defined inductively using matrix product with a constant matrix $\mathbf{c}$ and Hadamard product, proving the claim for these operations along with the base case proves the statement.

1. The base case is a linear form $u_i X_i + v_i Y_i$. By definition of φ_S,

   $\varphi_S(u_i X_i + v_i Y_i) = \big((u_i a_{j,j+1} + v_i b_{j,j+1})X_i + v_i b_{j,j+1} Y_i\big)$ when $s_j = i$ and is free of $b_{j,j+1}$ otherwise. Therefore,
   $\dfrac{\partial\, \mathrm{coef}\left(X_i, \varphi_S \circ (u_i X_i + v_i Y_i)\right)}{\partial b_{j,j+1}} = \frac{1}{b_{j,j+1}} \cdot \mathrm{coef}(Y_i, \varphi_S \circ (u_i X_i + v_i Y_i))$, when $s_j = i$ and 0 otherwise.
   $\dfrac{\partial\, \mathrm{coef}\left(Y_i, \varphi_S \circ (u_i X_i + v_i Y_i)\right)}{\partial b_{j,j+1}} = \frac{1}{b_{j,j+1}} \cdot \mathrm{coef}(Y_i, \varphi_S \circ (u_i X_i + v_i Y_i))$, when $s_j = i$ and 0 otherwise.

2. For $f := g \star h$ note that $\varphi_S(f) = \varphi_S(g) \star \varphi_S(h)$ since φ_S is a homomorphism.

   The definition of φ_S allows at most one common variable in $\varphi_S(g)$ and $\varphi_S(h)$ which is of the form $b_{i,i}$ but $\mathbf{b}^{\mathbf{e}}$ does not have it. Thus, $\dfrac{\partial\, \mathrm{coef}(w, \varphi_S \circ (g \star h))}{\partial \mathbf{b}^{\mathbf{e}}} = \dfrac{\partial\, \mathrm{coef}(w, \varphi_S \circ g)}{\partial \mathbf{b}^{\mathbf{e}_1}} \star \dfrac{\partial\, \mathrm{coef}(w, \varphi_S \circ h)}{\partial \mathbf{b}^{\mathbf{e}_2}}$, where $\mathbf{b}^{\mathbf{e}_1}$ and $\mathbf{b}^{\mathbf{e}_2}$ are over the variables occurring in $\varphi_S(g)$ and $\varphi_S(h)$ respectively. It is now obvious that if the claim holds for h and g, it holds for f.

3. If the statement is true for f, it is also true for $\mathbf{c} \cdot f$ because $\mathbf{c}$ is a matrix with constants and so, $\dfrac{\partial\, \mathrm{coef}(w, \varphi_S \circ \mathbf{c} \cdot f)}{\partial \mathbf{b}^{\mathbf{e}}} = \mathbf{c} \cdot \dfrac{\partial\, \mathrm{coef}(w, \varphi_S \circ f)}{\partial \mathbf{b}^{\mathbf{e}}}$.

□

We now prove the commutative concentration lemma for product depth-2 case. Recall that w_2 is the dimension of f_i in $C = \mathbf{c} \cdot f_1 \star \ldots \star f_d$, and w_1 is the dimension of u_j, v_j in $f_i = \mathbf{c} \cdot \prod_{j=n(i-1)+1}^{ni} (u_j X_j + v_j Y_j)$.

We define $\ell_0 := (\lceil \log w_1 \rceil + 1), \ell_1 := (\lceil \log w_2 \rceil + 1), \ell_2 := \ell_1 \ell_0$.

Lemma 4.6. *For $F := f_1 \star \ldots \star f_d$ let $\varphi := \varphi_{\ell_2}$. Then $F' := \varphi(F)$ is ℓ_2-concentrated. Consequently, $C' := \varphi(C)$ is ℓ_2-concentrated.*

Proof. The second statement follows from the first because any dependency in the coefficients of F' gives a dependency in the coefficients of C' by multiplication with the matrix $\mathbf{c}$.

Any rank or span in this proof is over K.

By the remark following Lemma 3.5, $\varphi(f_i)$ is ℓ_0-concentrated.

We first show $\mathrm{coef}(m, \varphi(F)) \in \mathsf{span}\{\mathrm{coef}(w, \varphi(F)) \mid w = \hat{w}_1 \dots \hat{w}_d, \forall i \in [d], \deg_{\mathbf{Y}} \hat{w}_i < \ell_0\}$, where $w = \hat{w}_1 \dots \hat{w}_d$ is block decomposition of w. Suppose block decomposition of $m = \hat{m}_1 \dots \hat{m}_d$ and $\deg_{\mathbf{Y}}(\hat{m}_i) \geq \ell_0$, then $\mathrm{coef}(\hat{m}_i, \varphi(f_i)) \in \mathsf{span}\{\mathrm{coef}(m', \varphi(f_i)) \mid \deg_{\mathbf{Y}}(m') < \ell_0, m' \in \mathrm{Supp}(f_i)\}$ by ℓ_0-concentration. Since $\mathrm{coef}(m, \varphi(F)) = \prod_{i=1}^{d} \mathrm{coef}(\hat{m}_i, \varphi(f_i))$, we can place $\mathrm{coef}(m, \varphi(F))$ in a coefficient space that corresponds to monomials having $\mathbf{Y}$-degree less than ℓ_0 in their i-th block. Repeating the argument with subsequent i proves the claim. For the lemma proof, it is thus sufficient to argue for the case of monomials that have $\mathbf{Y}$-degree less than ℓ_0 in each block.

We say that $\varphi(F)$ is ℓ_1-*block concentrated* if for any m, $\mathrm{coef}(m, \varphi(F)) \in \mathsf{span}\{\mathrm{coef}(w, \varphi(F)) \mid \mathrm{bdeg}_{\mathbf{Y}}(w) < \ell_1\}$. If $\varphi(F)$ is ℓ_1-block concentrated then $\varphi(F)$ is ℓ_2-concentrated since $\varphi(f_i)$ are already ℓ_0-concentrated.

We will now show ℓ_1 block concentration of $\varphi(F)$. For a monomial m with $\mathrm{bdeg}_{\mathbf{Y}}(m) = t (> \ell_1)$ and its block decomposition $m = \hat{m}_1 \dots \hat{m}_d$, let T be a set of block indices such that $|T| = \ell_1$ and $\forall i \in T$, $\deg_{\mathbf{Y}}(\hat{m}_i) \geq 1$. Let $F_T := \prod_{i \in T} f_i, m_T := \prod_{i \in T} \hat{m}_i$ and $m_{\bar{T}} := \prod_{i \in [d] \setminus T} \hat{m}_i$. If $S := \mathcal{Y}(m_T)$ then, $|S| \leq \ell_2$. Set α to α_S in $\varphi(F)$.

Finding a dependency. Let us denote the monomial with only $\mathbf{X}$ variables in f_i by $\hat{x}_i$. Consider a set of monomials $M := \prod_{i \in T} \{\hat{m}_i, \hat{x}_i\} \cdot m_{\bar{T}}$. All $w \in M \setminus m$ are such that $\mathrm{bdeg}_{\mathbf{Y}}(w) < t$. Since $|M| = 2^{\ell_1} > w_2$, there must exist a dependency among the coefficients corresponding to M in $\varphi_S(F)$. If the dependency does not have the coefficient of m, choose a coefficient corresponding to monomial the n such that it has the maximum $\mathrm{bdeg}_{\mathbf{Y}}(n)$. If $P := \mathcal{Y}(m) \setminus \mathcal{Y}(n)$ then $P \subseteq S$.

Let the dependency be:

$$\mathrm{coef}(n, \varphi_S \circ F) = \sum_{i=1}^{r} \beta_i \, \mathrm{coef}(n^{(i)}, \varphi_S \circ F) \text{ where } \beta_i \in K \setminus 0, n^{(i)} \in M$$

If the block decompositions are $n = \hat{n}_1 \dots \hat{n}_d$ and $n^{(i)} = \hat{n}_1^{(i)} \dots \hat{n}_d^{(i)}$, then, $n_T = \prod_{j \in T} \hat{n}_j = m_T = \prod_{j \in T} \hat{n}_j^{(i)} = n_T^{(i)}$. Also since $\mathrm{coef}(n, \varphi_S(F)) = \prod_{i=1}^{d} \mathrm{coef}(\hat{n}_i, \varphi_S(f_i))$, every term in the dependency has G_S as a factor where $G_S := \prod_{i \in [d] \setminus T} \mathrm{coef}(\hat{m}_i, \varphi_S \circ f_i)$. Let $G := \prod_{i \in [d] \setminus T} \mathrm{coef}(\hat{m}_i, \varphi \circ f_i)$. The dependency can also be written as

$$\mathrm{coef}(n_T, \varphi_S \circ F_T) \star G_S = \sum_{i=1}^{r} \beta_i \, \mathrm{coef}(n_T^{(i)}, \varphi_S \circ F_T) \star G_S$$

Deriving a dependency with coefficient of m. If $S = \{s_1, \dots, s_k\}$ and $\mathbf{b}^{\mathbf{e}} := \prod_{s_j \in S \cap P} b_{j,j+1}$ then, by Lemma 4.5

$$\frac{\partial \, \mathrm{coef}(n, \varphi_S \circ F)}{\partial \mathbf{b}^{\mathbf{e}}} = \beta \cdot \mathrm{coef}(m, \varphi_S \circ F)), \qquad \beta \in K \setminus 0$$

By choice of n, $\forall i \in [r], \exists j \in T$ such that $\hat{n}_j \neq \hat{x}_j$ but $\hat{n}_j^{(i)} = \hat{x}_j$. So P will exclude the indices in block j and by Lemma 4.5 if

$$\frac{\partial\, \mathrm{coef}(n^{(i)}, \varphi_S \circ F))}{\partial \mathbf{b}^{\mathbf{e}}} = \mathrm{coef}(w, \varphi_S \circ F),$$

then $\hat{w}_j = \hat{x}_j$. Hence, $\mathrm{bdeg}_{\mathbf{Y}}(w) \leq \mathrm{bdeg}_{\mathbf{Y}}(m)$. Since β_i may share variables with $\mathbf{b}^{\mathbf{e}}$,

$$\frac{\partial \beta_i\, \mathrm{coef}(n^{(i)}, \varphi_S \circ F))}{\partial \mathbf{b}^{\mathbf{e}}} \in \mathsf{span}\{\mathrm{coef}(w, \varphi_S \circ F)) \mid \mathrm{bdeg}_{\mathbf{Y}}(w) < \mathrm{bdeg}_{\mathbf{Y}}(m) = t\}$$

Differentiating the dependency by $\mathbf{b}^{\mathbf{e}}$ gives

$$\mathrm{coef}(m, \varphi_S \circ F) \in \mathsf{span}\{\mathrm{coef}(w, \varphi_S \circ F) \mid \mathrm{bdeg}_{\mathbf{Y}}(w) < t\}.$$

Since G_S does not share any variables with $\mathbf{b}^{\mathbf{e}}$ it is retained as a factor and so the above is the same as

$$\mathrm{coef}(\prod_{i\in T} \hat{m}_i, \varphi_S \circ F_T) \star G_S \in \mathsf{span}\{\mathrm{coef}(w, \varphi_S \circ F_T) \mid \mathrm{bdeg}_{\mathbf{Y}}(w) < \ell_1 \text{ and } w \in \mathrm{Supp}(F_T)\} \star G_S$$

Lifting a dependency on the coefficients of $\varphi_S(F)$ to a dependency on the coefficients of $\varphi(F)$. Since G_S is unit, we have proved that

$$\mathsf{span}\{\mathrm{coef}(w, \varphi_S \circ F_T) \mid \forall w \in \mathrm{Supp}(\varphi_S \circ F_T)\} = \mathsf{span}\{\mathrm{coef}(w, \varphi_S \circ F_T) \mid \mathrm{bdeg}_{\mathbf{Y}}(w) < \ell_1\}. \quad (4.7)$$

If we denote $\mathcal{W} := \mathsf{span}\{\mathrm{coef}(w, \varphi \circ F_T) \mid \forall w \in \mathrm{Supp}(\varphi \circ F_T)\}$ and $\mathcal{U} := \mathsf{span}\{\mathrm{coef}(w, \varphi \circ F_T) \mid \mathrm{bdeg}_{\mathbf{Y}}(w) < \ell_1\}$ then Equation 4.7 is can be restated as $\mathcal{U}|_{\alpha=\alpha_S} = \mathcal{W}|_{\alpha=\alpha_S}$.

Observe, $\mathrm{rk}\,\mathcal{U}|_{\alpha=\alpha_S} \leq \mathrm{rk}\,\mathcal{U} \leq \mathrm{rk}\,\mathcal{W} = \mathrm{rk}\,\mathcal{W}|_{\alpha=\alpha_S}$. The first inequality is true because the rank cannot increase by α substitution. Consider a matrix with columns as vectors of $\mathcal{U}|_{\alpha=\alpha_S}$. A nonzero minor here implies a corresponding nonzero minor in a matrix with columns as vectors of $\mathcal{U}$. The second inequality is justified as $\mathcal{U}$ is a subspace of $\mathcal{W}$ and Proposition 4.4 proves the third. Thus $\mathrm{rk}\,\mathcal{U} = \mathrm{rk}\,\mathcal{W}$ and $\mathcal{U} = \mathcal{W}$.

Therefore, now without the α substitution we have,

$$\mathsf{span}\{\mathrm{coef}(w, \varphi \circ F_T) \mid \forall w \in \mathrm{Supp}(\varphi \circ F_T)\} = \mathsf{span}\{\mathrm{coef}(w, \varphi \circ F_T) \mid \mathrm{bdeg}_{\mathbf{Y}}(w) < \ell_1\}. \quad (4.8)$$

Thus $\mathrm{coef}(\prod_{i\in T} \hat{m}_i, \varphi \circ F_T) \star G_S \in \mathsf{span}\{\mathrm{coef}(w, \varphi \circ F_T) \mid \mathrm{bdeg}_{\mathbf{Y}}(w) < \ell_1\} \star G_S$. Hadamard product with $G_S^{-1} \star G$, gives $\mathrm{coef}(m, \varphi \circ F) \in \mathsf{span}\{\mathrm{coef}(w, \varphi \circ F) \mid \mathrm{bdeg}_{\mathbf{Y}}(w) < t\}$.

Finally, induction on t gives ℓ_1-block concentration for $\varphi \circ F$. $\square$

4.2 Noncommutative results

We now show a relation between the coefficients of the noncommunicative model $\psi_n(C)$ and the corresponding commutative model $\varphi_n(C^{\text{com}})$ similar to the one shown in Proposition 3.10 for product depth-1.

Proposition 4.9. *Let* $F = \prod_{i=1}^{k} f_i$*. Interpret* $\operatorname{coef}(m, \psi_n \circ F)$ *as a block matrix with entries that are matrices of the same dimension as coefficients of* F*. Let* $C = \mathbf{c} \cdot F$*.*

1. *The* $(1, t+1)$*-th entry of the block matrix* $\operatorname{coef}(m, \psi_n \circ F)$ *is*

   $\operatorname{coef}(m, \psi_n \circ F)_{1,t+1} = \sum_S \operatorname{coef}(m^{\text{com}}, \varphi_{n,S}(F^{\text{com}}))$ *where summation is over all possible* $S \in \binom{[d]}{t}$*,* $d := \deg(F)$*.*

2. *The* (p, q)*-th entry of the block matrix* $\operatorname{coef}(m, \psi_n \circ F)$ *is*

   $\operatorname{coef}(m, \psi_n \circ F)_{p,q} = \sigma_{p-1} \circ \operatorname{coef}(m, \psi_n \circ F)_{1,q-p+1}$ *where* $\sigma_{p-1} : a_{i,j} \mapsto a_{i+p-1,j+p-1}$ *and* $b_{i,j} \mapsto b_{i+p-1,j+p-1}$*.*

3. *The* $(1, t+1)$*-th entry of the block matrix* $\operatorname{coef}(m, \psi_n \circ C)$ *is*

   $\operatorname{coef}(m, \psi_n \circ C)_{1,t+1} = \sum_S \operatorname{coef}(m^{\text{com}}, \varphi_{n,S}(C^{\text{com}}))$ *where summation is over all possible* $S \in \binom{[d]}{t}, d := \deg(C)$*.*

Proof. Proposition 3.10 proved the statement for f_i. For the first statement it is sufficient to show the claim for $f = g \star h$. Assume the claim hold for h and g by induction.

$\operatorname{coef}(m, \psi_n \circ f) = \operatorname{coef}(m_1, \psi_n \circ g) \star \operatorname{coef}(m_2, \psi_\ell \circ h)$ where $m = m_1 m_2, \deg(m_1) = \deg(g) = d_g, \deg(m_2) = \deg(h) = d_h$.

$\operatorname{coef}(m, \psi_n(f))_{1,t+1} = \sum_{k=1}^{t+1} \operatorname{coef}(m_1, \psi_n(g))_{1,k} \star \operatorname{coef}(m_2, \psi_n(h))_{k,t}$, interpreting $\psi_n(f), \psi_n(g)$ as block matrices with Hadamard multiplication for the entries.

Consider paths with $k-1$ jumps in the first $d_g := \deg(g)$ steps and $t-k+1$ jumps in the next $d_h := \deg(h)$ steps. Sum over all k is all possible paths of length $d_g + d_h$ with t jumps. This is formalized below.

$$\operatorname{coef}(m, \psi_n(f))_{1,t+1} = \sum_{k=1}^{t+1} \sum_{S_1 \in \binom{[d_g]}{k-1}} \operatorname{coef}(m_1^{\text{com}}, \varphi_{n,S_1}(g^{\text{com}})) \star \sigma_{k-1} \circ \sum_{S_2 \in \binom{[d_h]}{t-k+1}} \operatorname{coef}(m_2^{\text{com}}, \varphi_{n,S_2}(h^{\text{com}})).$$

Since τ_{d_g} which maps $X_i \to X_{i+d_d}$, $Y_i \mapsto Y_{i+d_g}$ is just renaming of variables,

$\operatorname{coef}(m_2^{\text{com}} \varphi_{n,S_2}(h^{\text{com}})) = \operatorname{coef}(\tau_{d_g} \circ m_2^{\text{com}}, \varphi_{n,S_2}(\tau_{d_g} \circ h^{\text{com}}))$. Let $T := S_1 \cup \{i + d_g \mid i \in S_2\}$, then

$$\operatorname{coef}(m_1^{\text{com}}, \varphi_{n,S_1}(g^{\text{com}})) \star \sigma_{k-1} \circ \operatorname{coef}(m_2^{\text{com}}, \varphi_{n,S_2}(h^{\text{com}})) = \operatorname{coef}(m^{\text{com}}, \varphi_{n,T}(g^{\text{com}} \star \tau_{d_g}(h^{\text{com}})))$$

Thus,

$$\begin{aligned}\operatorname{coef}(m, \psi_n(f))_{1,t+1} &= \sum_{S \in \binom{[d]}{t}} \operatorname{coef}(m^{\mathrm{com}}, \varphi_{n,S}\ (g^{\mathrm{com}} \star \tau_{d_g} \circ h^{\mathrm{com}})) \\ &= \sum_{S \in \binom{[d]}{t}} \operatorname{coef}(m^{\mathrm{com}}, \varphi_{n,S} \circ f^{\mathrm{com}})\end{aligned}$$

The same idea but with indices shifted proves the second statement.

For the last claim, let $\mathbf{c}$ be a matrix of constants, $f = \mathbf{c} \cdot g$, and the statement hold for g. By definition, $\operatorname{coef}(m, \psi_n \circ f)_{i,j} = \mathbf{c} \cdot \operatorname{coef}(m, \psi_n \circ g)_{i,j}$. For any S, $\mathbf{c} \cdot \operatorname{coef}(m^{\mathrm{com}}, \varphi_{n,S}(g^{\mathrm{com}})) = \operatorname{coef}(m^{\mathrm{com}}, \varphi_{n,S}(\mathbf{c} \cdot g^{\mathrm{com}})) = \operatorname{coef}(m^{\mathrm{com}}, \varphi_{n,S}(f^{\mathrm{com}}))$ and this proves the claim. □

4.3 Concentration in product-depth 2

The next lemma shows that to check the nonzeroness of the noncommutative circuit C, it is sufficient to check the nonzeroness of monomials of low y-degree in $\psi_n(C)$ for appropriate n.

Lemma 4.10. *Let $C = \mathbf{c} \cdot f_1 \star \ldots \star f_d$ be a noncommutative polynomial with $f_i \in H_{w_2}(\mathbb{F})$, $f_i = \mathbf{c}_1 \cdot \prod_{j=1}^{d}(u_{ij} \cdot x + v_{ij} \cdot y)$, $u_{ij}, v_{ij} \in H_{w_1}(\mathbb{F})$. Define $\ell_0 := \lceil \log w_1 \rceil + 1, \ell_1 := \lceil \log w_2 \rceil + 1, \ell_2 := \ell_0 \ell_1$. Then, $C \equiv 0 \iff \forall m, \deg_y(m) < \ell_2, \operatorname{coef}(m, \psi_{\ell_2}(C)) = \mathbf{0}$.*

Remark 4.11. For the proof it is crucial to have $\mathbf{c} \in \mathbb{F}^w$ (i.e., each coordinate is a constant, in $\mathbb{F}$, free of ψ_{ℓ_2} variables) otherwise the path argument using S breaks down.

Proof. Only the backward direction is non-trivial. Assume $\forall m, \deg_y(m) < \ell_2, \operatorname{coef}(m, \psi_{\ell_2}(C)) = \mathbf{0}$. In particular, for all such coefficients, the $(1, \ell_2 + 1)$-th entry is zero.

The third claim of Proposition 4.9 shows, $\operatorname{coef}(m, \psi_{\ell_2}(C))_{1,\ell_2+1} = \sum_S \operatorname{coef}(m^{\mathrm{com}}, \varphi_{\ell_2,S}(C^{\mathrm{com}}))$ where summation is over all possible $S \in \binom{[d]}{\ell_2}$.

Since for different sets S, $\operatorname{coef}(m^{\mathrm{com}}, \varphi_{\ell_2,S}(C^{\mathrm{com}}))$ have disjoint support (Proposition 4.3), $\operatorname{coef}(m^{\mathrm{com}}, \varphi_{\ell_2,S}(C^{\mathrm{com}})) = 0$ for all subsets S of $[d]$ of size ℓ_2.

Thus we have $\forall m^{\mathrm{com}}, \deg_y(m^{\mathrm{com}}) < \ell_2, \forall S \in \binom{[d]}{\ell_2}, \operatorname{coef}(m^{\mathrm{com}}, \varphi_{\ell_2,S}(C^{\mathrm{com}})) = 0$. This together with Lemma 4.6 which proves ℓ_2-concentration of $\varphi_{\ell_2}(C^{\mathrm{com}})$ implies $\varphi_{\ell_2}(C^{\mathrm{com}}) \equiv 0$ (Lemma 3.8). As φ_{ℓ_2} is an invertible transformation, $C^{\mathrm{com}} \equiv 0$. We finally arrive at the conclusion in the lemma using $C^{\mathrm{com}} \equiv 0 \iff C \equiv 0$ (Proposition 4.2). □

Recall a depth-2 $\sum\prod\sum\prod\sum$ circuit C over $H_{k^0}(\mathbb{F})$ with product gate fan-in d and sum gate fan-in k can be expressed as $C = \mathbf{c}_1 \cdot f_1 \star \ldots \star f_d$, where each $f_i = \mathbf{c} \cdot \prod_{j=1}^{d}(U_{ij} \cdot x + V_{ij} \cdot y)$, $U_{ij}, V_{ij} \in H_{k^2}(\mathbb{F}), \mathbf{c} \in \mathbb{F}^{k \times k^2}$. The degree of the polynomial computed by C is d^2. If the size of the circuit is s, the degree is at most $\mathrm{poly}(s)$.

Proof of Theorem 1 for bivariate polynomial computed by a $\sum\prod\sum\prod\sum$ circuit.

Lemma 4.10 shows that for $\ell_0 = O(\log k^2)$, $\ell_1 = O(\log k)$ and so $\ell_2 = \ell_0 \ell_1 = O(\log^2 k)$, it is enough to check the nonzeroness of the coefficients corresponding to monomials with y-degree less than ℓ_2 in $\psi_{\ell_2}(C(x,y))$. This can be done by substituting x and y with matrices $M_{\mathbf{x},\ell_2}$ and $M_{\mathbf{y},\ell_2}$ as shown in Lemma 3.9.

Therefore, if $C(x,y) \not\equiv 0$ then $C(A \otimes M_{\mathbf{x},\ell_2}, B \otimes (M_{\mathbf{x},\ell_2} + M_{\mathbf{y},\ell_2})) \not\equiv 0$. Here A, B as well as $M_{\mathbf{x},\ell_2}, M_{\mathbf{y},\ell_2}$ are $(\ell_2+1) \times (\ell_2+1)$ matrices. The final map is a substitution of x and y by symbolic matrices of dimension $O(\log^4 k) = O(\log^4 s)(\because k \leq s)$. The output is a matrix with entries that are polynomials of degree $2d^2$ in the commutative variables $\{a_{ij}, b_{ij}, x_{ij}, y_{ij} \mid i,j \in [\ell_2+1]\}$. The polynomial identity Lemma 2.1 over an extension field of dimension $O(\log s)$ for a nonzero entry of the matrix gives the PIT algorithm. In conclusion, the blackbox randomized PIT algorithm is to substitute x, y with matrices of dimension $O(\log^4 s)$ with random entries from an extension field of degree $O(\log s)$. $\square$

5 Proof for constant depth

The proof follows the same template as the previous two sections. Given a homogeneous constant depth noncommutative circuit C, first prove concentration in the corresponding commutative circuit after applying the appropriate subset cover map, φ_n. Then establish a connection between $\psi_n(C)$ and the commutative model. Using the concentration in commutative model, argue for the nonzeroness in low y-degree coefficients in $\psi_n(C)$. Finally, reduce the check of nonzeroness of these fewer coefficients to a commutative PIT check by substituting x, y with matrices.

Note that all the proofs for the product depth-2 result were based on induction. The corresponding results in product depth-1 proved the base cases. The results for product depth-2 showed that the desired properties continue to hold with Hadamard product and matrix multiplication with a matrix of constants. Hence if the claims are true for depth-$(\Delta - 1)$, then they are also true for depth-Δ.

5.1 Finishing up constant depth

Proof of Theorem 1. Let $C' \in \mathbb{F}\langle \mathbf{z} \rangle$ be a homogeneous depth-Δ' circuit of size s'. There exists a corresponding bivariate homogeneous circuit $C \in \mathbb{F}\langle x, y \rangle$ of depth at most $\Delta := \Delta' + 1$ and size $s := O(s' + n^2)$ such that $C' \equiv 0$ if and only if $C \equiv 0$ as shown in Proposition 3.1. The construction also showed that given blackbox access to C', we can create from it an efficient blackbox access for C. Hence it is sufficient to give a randomized blackbox PIT for bivariate C in time $\text{poly}(\log^{\Delta} s)$ as it proves the result for C'.

For the bivariate circuit C if the output gate is a product gate, then PIT on factors gives PIT for the product because in a matrix algebra nonzeroness implies invertibility by Theorem 3. Hence if the factors are invertible their product cannot be zero. It can now be assumed that the output gate is a sum gate.

Let C, with product fanin d and sum fanin k, be $C := \sum_{i=1}^{k}\prod_{j=1}^{d} g_{i,j}$ where $g_{i,j}$ are depth $\Delta-2$ circuits over $H_w(\mathbb{F})$. Then $C = \mathbf{c}\cdot\prod_{j=1}^{d}\left[g_{1j}, g_{2j}, \dots, g_{kj}\right]^T$ for appropriate $\mathbf{c}$. Let $\ell_{\Delta,w}$ be a parameter such that $\varphi_{\ell_{\Delta,w}}(C^{\mathrm{com}})$ is $\ell_{\Delta,w}$-concentrated.

Let $f_j = (g_{1j}, g_{2j}, \dots, g_{kj})^T$, then $C = \mathbf{c}\cdot f_1 \star f_2 \star \dots \star f_d$. Each f_j is $\Delta-2$ depth circuit over $H_{kw}(\mathbb{F})$. Let $\ell_0 := \ell_{\Delta-2,kw}$ and assume by induction that $\forall i \in [d], \varphi_{\ell_0}(f_i^{\mathrm{com}})$ are ℓ_0-concentrated.

The proof of commutative concentration (Lemma 4.6) shows that for $\ell_1 := \lceil \log kw\rceil + 1$, $\varphi_{\ell_1\ell_0}(C^{\mathrm{com}})$ will be $\ell_1\ell_0$-concentrated. The same proof works because it only required that $\varphi_{\ell_0}\circ f_i^{\mathrm{com}}$ are ℓ_0-concentrated. The results of Proposition 4.4 and Lemma 4.5 referred there, were already proved using the inductive definition of C^{com}. Thus $\ell_{\Delta,w} := \ell_1\ell_0 = (\lceil \log kw\rceil + 1)\ell_{\Delta-2,kw}$ and $\ell_{3,w} = \lceil \log w\rceil + 1$ (Lemma 3.5) and $\ell_{2,w} = O(\log k)$ ([AJMR19] for sparsity k). Solving the recursive system, we get $\ell_{\Delta,w} = O(\log^{\Delta/2} kw)$. Since the given circuit is over $\mathbb{F} = H_1(\mathbb{F})$, setting w to 1 we get $\ell_{\Delta,1} = O(\log^{\Delta/2} k)$. Denote this by ℓ in the rest of the proof.

Next, the relation between noncommutative circuit $\psi_n(C)$ and commutative circuit $\varphi_n(C^{\mathrm{com}})$ established in Proposition 4.9 was also proved for the inductive definition.

Finally, an analogous Lemma 4.10 is also true since it only required the commutative concentration of $\varphi_\ell(C^{\mathrm{com}})$ and the relation above. As a result we only need to check the coefficients of $\psi_\ell(C)$ that correspond to monomials with y-degree at most ℓ. This can be done by substituting x and y with matrices $M_{\mathbf{x},\ell}$ and $M_{\mathbf{y},\ell}$ as shown in Lemma 3.9.

Therefore, if $C(x,y) \not\equiv 0$ then $C(A\otimes M_{\mathbf{x},\ell}, B\otimes(M_{\mathbf{x},\ell}+M_{\mathbf{y},\ell})) \not\equiv 0$. Here A, B as well as $M_{\mathbf{x},\ell}, M_{\mathbf{y},\ell}$ are $(\ell+1)\times(\ell+1)$ matrices. The final map is a substitution of x and y by symbolic matrices of dimension $O(\log^{\Delta} k) = O(\log^{\Delta} s)(\because k \le s)$. The output is a matrix with entries that are polynomials of degree $\mathrm{poly}(s)$ in the commutative variables $\{a_{ij}, b_{ij}, x_{ij}, y_{ij} \mid i,j \in [\ell+1]\}$. The polynomial identity Lemma 2.1 over an extension field of dimension $O(\log s)$ for a nonzero entry of the matrix gives the PIT algorithm. In conclusion, the blackbox randomized PIT algorithm is to substitute x, y with matrices of dimension $O(\log^{\Delta} s)$ with random entries from an extension field of degree $O(\log s)$. $\square$

6 Lower bound for matrix identities

Proof of Theorem 2. Let a nonzero polynomials $f(z_1,\dots,z_n)$ be represented by a depth-$(\Delta-1)$ homogeneous circuit of size s. By Theorem 1 we know that f cannot be an identity for $\mathcal{M}_{O(\log^{\Delta} s)}(L)$ as there is at least one tuple of matrices $(M_1,\dots,M_n)$ with $M_i \in \mathcal{M}_{O(\log^{\Delta} s)}(L)$ such that $f(M_1,\dots,M_n)$ is nonzero. Let $d := O(\log^{\Delta} s)$, then $s = \exp(\Omega(d^{1/\Delta}))$ and so any f of size s, depth $\Delta-1$ cannot be an identity for $\mathcal{M}_d(L)$. $\square$

7 Conclusion

The idea of low-support rank concentration has now been shown to be useful in the noncommutative setting, with a significant adaptation. The natural next questions are: If the techniques here can be adapted to give an efficient PIT for homogeneous noncommutative circuits of unrestricted depth and subsequently a PIT for general noncommutative circuits.

Acknowledgements. N.S. thanks the funding support of DST-SERB (JCB/2022/57), and the Chair of N. Rama Rao (2019–) of the Department of CSE, IIT Kanpur. The authors thank Anagha Gokul for the many discussions in the initial days of research on this problem [Gok25].

References

[ACM24] V. Arvind, Abhranil Chatterjee, and Partha Mukhopadhyay. Black-Box Identity Testing of Noncommutative Rational Formulas in Deterministic Quasipolynomial Time. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, page 106–117, New York, NY, USA, 2024. Association for Computing Machinery. 3

[AJMR19] Vikraman Arvind, Pushkar S. Joglekar, Partha Mukhopadhyay, and S. Raja. Randomized Polynomial-Time Identity Testing for Noncommutative Circuits. *Theory of Computing*, 15(7):1–36, 2019. 1, 3, 5, 8, 13, 26

[AKS04] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. PRIMES is in P. *Annals of Mathematics*, pages 781–793, 2004. 2

[AL50] Avraham Shimshon Amitsur and Jakob Levitzki. Minimal identities for algebras. *Proceedings of the American Mathematical Society*, 1(4):449–463, 1950. 2

[Ami55] Shimshon Avraham Amitsur. The T-ideals of the free ring. *Journal of the London Mathematical Society*, s1-30(4):470–475, 1955. 8

[Ami72] Shimshon Avraham Amitsur. On central division algebras. *Israel Journal of Mathematics*, 12(4):408–420, 1972. 8

[Ami74] Shimshon Avraham Amitsur. The generic division rings. *Israel Journal of Mathematics*, 17(2):241–247, 1974. 8

[ASS13] Manindra Agrawal, Chandan Saha, and Nitin Saxena. Quasi-polynomial hitting-set for set-depth-Δ formulas. In *Proceedings of the 45th ACM Symposium on Theory of Computing, STOC 2013, Palo Alto, CA, USA, June 1-4, 2013*, pages 321–330, 2013. 4, 5

[BW05] A. Bogdanov and H. Wee. More on noncommutative polynomial identity testing. In *20th Annual IEEE Conference on Computational Complexity (CCC'05)*, pages 92–99, 2005. 1, 2, 4

[CGG+26] Abhranil Chatterjee, Sumanta Ghosh, Rohit Gurjar, Roshan Raj, and Thomas Thierauf. Bipartite Matching is in NC. *Electronic Colloquium on Computational Complexity (ECCC)*, 2026. 2

[DL78] Richard A. DeMillo and Richard J. Lipton. A probabilistic remark on algebraic program testing. *Information Processing Letters*, 7(4):193–195, 1978. 6

[DM17] Harm Derksen and Visu Makam. Polynomial degree bounds for matrix semi-invariants. *Advances in Mathematics*, 310:44–63, 2017. 3

[FGT19] Stephen A. Fenner, Rohit Gurjar, and Thomas Thierauf. A deterministic parallel algorithm for bipartite perfect matching. *Commun. ACM*, 62(3):109–115, 2019. 2

[GGOW16] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. A Deterministic Polynomial Time Algorithm for Non-commutative Rational Identity Testing. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 109–117, 2016. 3

[Gok25] Anagha Gokul. Polynomial identity testing of non-commutative circuits. Master's thesis, CSE, IIT Kanpur and BITS Pilani-Hyderabad, 2025. 27

[HH21] Masaki Hamada and Hiroshi Hirai. Computing the nc-Rank via Discrete Convex Optimization on CAT(0) Spaces. *SIAM Journal on Applied Algebra and Geometry*, 5(3):455–478, 2021. Pre-print available at `arXiv:https://doi.org/10.1137/20M138836X`. 3

[Hya77] Laurent Hyafil. The Power of Commutativity. In *Proceedings of the 18th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1977)*, pages 171–174. IEEE, 1977. 2

[IQS18] Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. Constructive non-commutative rank computation is in deterministic polynomial time. *Computational Complexity*, 27(4):561–593, 2018. 3

[Kos09] Bertram Kostant. A theorem of Frobenius, a theorem of Amitsur-Levitski and cohomology theory. In *Collected Papers: Volume I 1955-1966*, pages 64–91. Springer, 2009. 2

[Mak18] Viswambhara Makam. *Invariant Theory, Tensors and Computational Complexity*. Ph.d. dissertation, University of Michigan, 2018. Theorem 4.2.8. 8

[Nis91] N. Nisan. Lower Bounds for Non-Commutative Computation. In *Proc. 23rd ACM Symp. on Theory of Computing*, pages 410–418, 1991. Pre-print available at `arXiv:stoc:103418.103462`. 2, 3

[Ore22] Øystein Ore. Über höhere Kongruenzen. *Norsk Mat. Forenings Skrifter*, 1(7):15, 1922. 6

[Sax09] Nitin Saxena. Progress on Polynomial Identity Testing. *Bull. EATCS*, 99:49–79, 2009. 2

[Sax14] Nitin Saxena. Progress on polynomial identity testing-II. In *Perspectives in Computational Complexity*, pages 131–146. Springer, 2014. 2

[SBR25] G. V. Sumukha Bharadwaj and S. Raja. Randomized Black-Box PIT for Small Depth +-Regular Non-Commutative Circuits. In *45th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2025)*, volume 360 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 51:1–51:16, 2025. 1, 3

[Sch80] Jacob T. Schwartz. Fast Probabilistic Algorithms for Verification of Polynomial Identities. *Journal of the ACM (JACM)*, 27(4):701–717, 1980. 6

[SV85] Sven Skyum and Leslie Valiant. A complexity theory based on Boolean algebra. *Journal of the ACM*, 32(2):484–502, 1985. 2

[Val79] Leslie G. Valiant. Completeness Classes in Algebra. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing (STOC 1979)*, pages 249–261. ACM, 1979. 2

[Zip79] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Symbolic and Algebraic Computation, EUROSAM '79*, volume 72 of *Lecture Notes in Computer Science*, pages 216–226. Springer, 1979. 6