DEPARTMENT OF MATHEMATICS & STATISTICS

MASTER'S THESIS

PIT and separation between low-variate Read Once ABP classes

Author:
Sagar Arora

Supervisor:
Dr. Nitin Saxena
Dr. Arnab Hazra

A thesis submitted in fulfillment of the requirements for the degree of Master Of Science

to the

Indian Institute Of technology Kanpur

November, 2022

Declaration of Authorship

I, Sagar Arora, declare that this thesis titled, "PIT and separation between low-variate Read Once ABP classes" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Abstract

PIT and separation between low-variate Read Once ABP classes

authored by by Sagar Arora

supervised by Dr. Nitin Saxena, Dr. Arnab Hazra

The algebraic model of computation, due to its simplicity, elegance and abstract connections to many open problems has attracted a large amount of research in the last few decades. Algebraic circuits and Algebraic Branching Programs are the fundamental models for computing polynomials.The framework of algebraic computations has a beautiful interplay between well known problems in mathematics like identity testing polynomial equivalence, primality testing, graph isomorphisms and polynomial factoring.
Identity Testing is the problem of checking efficiently whether a given arithmetic circuit $\equiv 0$. We already have a polynomial time randomised algorithm for PIT. However, designing an efficient deterministic identity test has been a long-standing open question. Identity Testing is also a good candidate to study the derandomization questions and their connections.
The motivation of this thesis is to study closely the relationship between certain subclasses of the computational model of *Read once Oblivious Branching Programs* - commutative ROABPs and diagonal ROABPs. And in this attempt, devise an efficient Polynomial Identity Testing algorithms for ROABPs where the number of variables is low (logarithmic with respect to the circuit size). We show that if the dimension of partial derivatives and Waring rank of polynomials is similar upto polynomial factors, then the model of diagonal ROABPs efficiently captures the more descriptive model of commutative ROABPs.
We also study Blackbox PIT of log-variate ROABP by devising efficient polynomial shifting maps. The techniques build upon by investigating the structure of Newton polytopes of the polynomials and constructing weight assignments in a fashion that the coefficient space is efficiently captured by a low number of monomials.

Acknowledgements

I would like to begin by thanking Prof. Nitin Saxena for his encouragement throughout the study and sharing his key insights which always motivated me to develop a better understanding of the problem.

I would like to make a special mention of gratitude to Prof. Arnab Hazra for accepting to be my departmental supervisor and allowing me the opportunity to work on this problem even though this domain is not his primary area of research. I would like to thank Prateek Dwivedi, for his continued assistance during the course of this study; for being available for discussions and clearing my stupid doubts.

Lastly, I would like confer my gratitude to my family, friends and my football team for their unwavering belief in me.

Contents

List of Figures

Chapter 1

Introduction

1.0.1 Identity Testing

It is a well known fact that a d - degree polynomial over a field, can have at most d roots. Thus, there is a simple test for the non-zeroness of the polynomial - evaluate at $d+1$ distinct points.

Schwartz Zippel Lemma

Theorem 1.1. Zippel, 1979 Let $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be a degree d, n - variate polynomial over a field $\mathbb{F}$. Let $S \subseteq \mathbb{F}$ be a set of size $> d$. Then

$$Pr[f(a_1, a_2, \ldots, a_n) \neq 0] \geq 1 - \frac{d}{|S|}$$

, where a_i is chosen uniformly randomly from S for each i, independently.

In case, we have the knowledge of upper bound on the individual degree of the polynomial f, the bound can be improved

$$Pr[f(a_1, a_2, \ldots, a_n) \neq 0] \geq (1 - \frac{d}{|S|})^n$$

This test needs the field size to be large enough. In case of finite fields, one can also work with appropriate field extensions.

Hitting set : A set of points H in the underlying domain $\mathbb{F}^n$ for a class $\mathcal{C}$ of n variate polynomials is a hitting set if for any non-zero polynomial f in $\mathcal{C}$, there exists a point in H where f evaluates to non-zero.

The main question thus, is how to generate small hitting sets and that too efficiently ?

Let $\mathcal{C}(n, d, s)$ be the set of algebraic circuits of size $\leq s$ computing polynomials in $\mathbb{F}[x_1, \ldots, x_n]$ of degree $\leq d$. Let C be a particular such circuit (which belongs to some class $\mathbf{C} \subset \mathcal{C}$), computing a polynomial f_C.

The problem of **PIT** asks whether $f_C \equiv 0$

The term "identity-testing" is an acknowledgment of the fact the we are trying to verify whether the model is a computational framework of some (non-trivial) algebraic identity like $(a+b)^2 - a^2 - b^2 \,|a, b, \in \mathbb{F}_2$ or $(a+b)(a-b) - a^2 + b^2$ over $\mathbb{C}$.

Randomized PIT : [Schwartz Zippel Lemma] introduces us to a probabilistic yet fast algorithm - it works in polynomial time even for polynomials with exponential degrees. In that case, the set S will be of exponential size, but a random element from S would need just $O(n)$ bits.

Deterministic PIT : Is there an efficient (poly (n, d, w))- time deterministic algorithm, which when given as input any n-variate, d-degree, size s circuit $mathcalC$ determines if $mathcalC \equiv 0$

Blackbox PIT : Another interesting way to look at the **[Zippel, 1979]** randomized algorithm is that it only need *evaluation points* as input to the circuit being tested; and does not need knowledge about the structure of the circuit. These tests, hence are referred to as "Blackbox PIT algorithms"

For blackbox PIT, we generally have to allow the evaluation points to be from the field extensions. For instance, consider a univariate $f \in \mathbb{F}_2$ where $f(x)x^2 - x$. Here, over all the elements of the field $f(x) = 0$. So to obtain an evaluation point such that f evaluates to a non-zero value we must go to the extensions of $\mathbb{F}_2$

The randomised PIT algorithm due to **[Zippel, 1979]** is a black-box PIT algorith as it does not require the knowledge of the structure of the computational model to carry out the tests. If, however, we try to derandomize the algorithm trivially, we get an exponential size hitting set computable in the same time complexxity. In particular, **[Heintz and Schnorr, 1980]** annd later **Agrawal, 2005** that constructing hitting sets for arithmetic circuits in polynomial time imply exponential size lower bounds for arithmetic circuits.
PIT has applications in designing various algorithms as well as proving various circuit lower bounds. With this natural measure for the complexity of polynomials at hand, To get a flavor about the prowess of the computational model that we would be dealing with in this study, we state the following chain of reductions between the arithmetic computational models.

$$\begin{aligned}\text{constant-depth arithmetic circuits} \leq_p \text{constant width ABP} \\ =_p \text{Formulas} \leq_p ABP \leq_p \text{Arithmetic circuits}\end{aligned} \tag{1.1}$$

$$\Sigma \bigwedge \Sigma \subsetneq diagROABp \subseteq commROABP \subseteq ROABP[\forall] \subseteq ROABP[\exists] \tag{1.2}$$

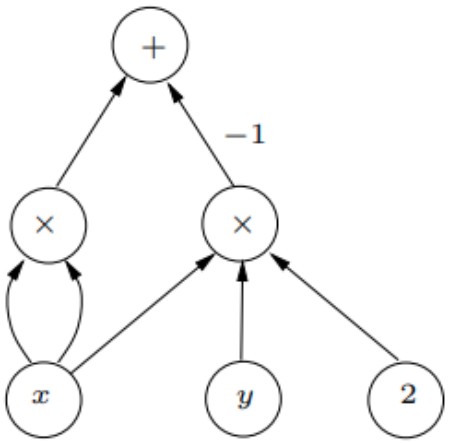

FIGURE 1.1: Arithmetic circuit computing $x^2 - 2xy$

Chapter 2

Preliminaries

2.1 Notations

2.1.1 Polynomials

- Throughout the report, $[n]$ denotes the set $\{1, 2, \ldots, n\}$.
- By $\mathbf{x}$ we denote the set of variables $\{x_1, x_2, \ldots, x_n\}$. For a set of n variables $\mathbf{x}$ and for an exponent $\mathbf{e} = (e_1, e_2, \ldots, e_n) \in \mathbb{W}^n$ and $\mathbf{x}^\mathbf{e}$ will denote the monomial $\prod_{i=1}^{n} x_i^{e_i}$.
- The support of a monomial $\mathbf{x}^\mathbf{e}$, denoted by *Supp(*$\mathbf{e}$*)*, is the set of variables appearing in that monomial, i.e. $supp(\mathbf{e}) = \{x_i \mid \mathbf{e}_i > 0, i \in [n]\}$.
- The support size of a monomial is the cardinality of its support, also denoted by $supp(\mathbf{e})$. This definition of support can be naturally extended to polynomials as collection of all the monomials of the polynomial whose support size is greater than zero.
- For a polynomial $f(\mathbf{x})$, the coefficient of a monomial $\mathbf{x}^\mathbf{e}$ is denoted by $coeff_f(\mathbf{x}^\mathbf{e})$. Also, as each monomial is uniquely identified by its exponent vector, there is an abuse of notations to define the support, coefficient and partial derivative of the monomials.
- For each monomial $\mathbf{e}$, e_i is said to be its degree in the i-th variable x_i and $\sum_i e_i$ is defined to be its (overall) degree.
- Naturally, extending this definition for a polynomial, we have
  - $deg_{x_i}(f) = max\{e_i \mid coeff_f(\mathbf{x}^\mathbf{e}) \neq 0, \mathbf{e} \in \mathbb{W}^n\}$
  - $deg(f) = max\{\sum_{i=1}^{n} e_i \mid coeff_f(\mathbf{x}^\mathbf{e}) \neq 0, \mathbf{e} \in \mathbb{W}^n$
  - the individual degree of f denoted by $indv - deg(f) = max\{deg_{x_i}(f) \mid i \in [n]\}$

2.1.2 Matrix Algebra

Throughout the report, as a bridge between the computational structure and the corresponding polynomial, we are going to deal with $HadamardAlgebra$ and $MatrixAlgebra$.
$\mathbb{F}^{m \times n}$ represents the set of all $m \times n$ matrices over the field $\mathbb{F}$.

We view a matrix with polynomial entries, as a polynomial with matrix coefficients. For instance,

$$f(x,y) = \begin{bmatrix} 1+x+x^2 & x+y \\ y-xy & 1+xy \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot 1 + \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \cdot x + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot x^2 + \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix} \cdot xy + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot y$$

2.1.3 Monomial Ordering

We fix a total ordering on the monomials that respects division/multiplication and fix a consistent way of comparing multivariate monomials by *lexicographic ordering*. This type of ordering is crucial for representing polynomial residues after quotienting in the multivariate setting. $\mathbf{t}^{\mathbf{e}} \prec \mathbf{t}^{\mathbf{e}'}$ if the smallest $i \in \mathbb{N}$ with $e_i \neq e'_i$ is such that $e_i < e'_i$

2.1.4 Hasse derivative space

For monomials $\mathbf{a}, \mathbf{e}$, we define the partial derivative of $\mathbf{x}^{\mathbf{a}}$ corresponding to $\mathbf{x}^{\mathbf{e}}$ as $\partial_{\mathbf{e}} \mathbf{x}^{\mathbf{a}} = \dfrac{|\partial^{\mathbf{e}}|}{\partial x_1^{e_1} \dots \partial x_n^{e_n}} \mathbf{x}^{\mathbf{a}}$

This definition can naturally be extended to partial derivatives of polynomials

$$\partial_{\mathbf{e}} f = \frac{|\partial^{\mathbf{e}}|}{\partial x_1^{e_1} \dots \partial x_n^{e_n} f}$$

Derivative operators :A derivative operator $\mathcal{D}$ on $\mathbb{F}[x_1, \dots, x_r]$ is an $\mathbb{F}$ - linear operator of finitely many partial derivative of the form $\partial_e : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{x}], where \mathbf{e} \in \mathbb{W}^r$

$$\mathcal{D} = \sum_{\mathbf{e}} \xi_{\mathbf{e}} \partial_{\mathbf{e}}$$

There is a one to one correspondence between a polynomials and the derivative operator space.

Any polynomial $g(\mathbf{x})$ naturally defines a derivative operator $\mathcal{D}_g : \sum_{\mathbf{e} \in supp(g)} coeff_g(\mathbf{e}) \cdot \partial_{\mathbf{e}}$

For any polynomial $g(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ and the corresponding derivative operator $\mathcal{D}_g$, we define the closure $\overline{\mathcal{D}_g}$ of $\mathcal{D}_g$ as

$$\overline{\mathcal{D}_g} := \{\mathcal{D}_{\partial_{\mathbf{e}}(g)} \mid \mathbf{e} \in \mathbb{W}^n, \partial_{\mathbf{e}}(g) \neq 0\} \tag{2.1}$$

Note that the closure of the derivative operator of polynomial g consists of all the *down-shifted* monomials of g, i.e. $\{\mathbf{e}' \mid \mathbf{e}' \prec \mathbf{e}\, \partial_{\mathbf{e}'}(g) \neq= 0\}$

Let $V(J)$ denote the variety of the ideal J. Define $\overline{\mathcal{D}(J)}$ the closure of the derivative operator space for J as

$$\overline{\mathcal{D}(J)} = \{\mathcal{D} \in \mathbb{F}[\mathbf{x}] \mid \mathcal{D}_{\mathbf{e}}(g)\big|_{\vec{\theta}_0}, \vec{\theta} \in V(J)\} \tag{2.2}$$

For each such point in the variety of an ideal, we can thus, construct a closed vector space of derivative operators such that every polynomial in the ideal evaluates to 0.

For an ideal J with finite variety (say τ) , there exist closed spaces of derivative operators $\mathcal{D}_1, \dots, \mathcal{D}_u$ of dimensions $r_1, \dots, r_u$. For any polynomial $g \in \mathbb{F}[\mathbf{x}]$, we have $d \in J$ such that $\Leftrightarrow \forall u \in [\tau], \mathcal{D}_v(g)(\overline{\theta}) = 0$

2.2 Models Of computation

2.2.1 Arithmetic Circuits

An arithmetic circuit is a natural computation model for polynomials. They are directed acyclic graph a unique sink (the output gate). The leaves are source vertices which take as input either a constant from the underlying or a variable from the support set $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$. Every internal node is labelled either by $+$ (addition gate) or $\times$ (multiplication gate). Every edge of this DAG carries weights that are elements from the underlying field $\mathbb{F}$.

Computation of the polynomial : Every edge collects the polynomial computed at its tail node, scales it up the weight on the edge and sends it to the head node. An addition gate computes the sum of inbound polynomials and every multiplication gate computes the product of all inbound polynomials. The formal polynomial computed at the sink is referred to as the polynomial computed by the circuit.

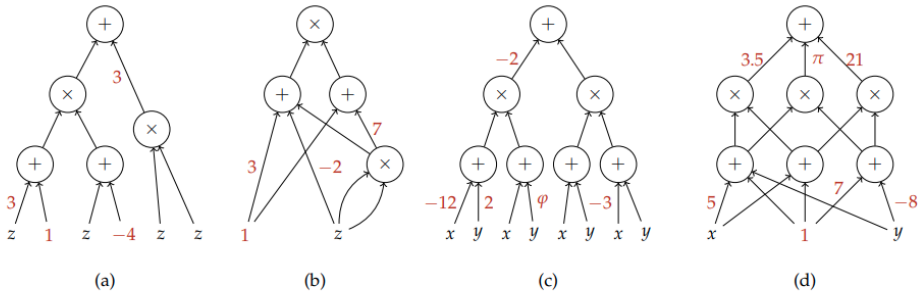

FIGURE 2.1: Arithmetic circuits and formulas

Characteristics :

- **depth -**The length of the longest path in the circuit from a leaf gate to an output gate
- **size -** The number of edges
- **degree -** The syntactic degree (note this this formal degree might not be the same as the actual degree of the polynomial)

Algebraic Formula It is an algebraic circuit whose underlying DAG is a tree

2.2.2 Depth-3 Diagonal circuits

Depth 3 Diagonal circuits (denoted by $\Sigma \cap \Sigma$ compute polynomials of the form

$$f(x) = \sum_{i=1}^{s} l_i^{d_i}$$

where each l_i is a linear polynomial over the underlying field $\mathbb{F}$. The rank of a depth 3 diagonal circuit (denoted by $rk(C)$), is the dimension of the subspace (over $\mathbb{F}$) generated by f_i. This rank can be shown to be equal or one less than the dimension of the subspace generated by l_i's.

subsectionAlgebraic Branching Programs **[Nisan, 1991]** An ABP over is a directed acyclic graph , with a unique source vertex u and a sink vertex t, and the edges have polynomials as their weights. The polynomials on the edges are 'simple', in a sense that they are linear over the support set of variables.

Computation : For an edge **e** denote its weight by $wt(\mathbf{e})$. Now consider a path p from any vertex a to b, The weight of this path denoted b $wt(p)$ is defined as the product of the edge weights of all the edges along this path, i.e. $wt(p) = \prod_{\mathbf{e}inp} wt(\mathbf{e})$. The formal polynomial computed by the ABP, then is $\sum_{p \in paths(u,t)} wt(p)$

Layered ABP Characteristics :
We can redefine an ABP as a directed graph layered with vertex set V and edge-set E such that $E = E_1 \sqcup E_2 \ldots \sqcup E_d$ where $E_i \subseteq V_{i-1} \times V_i$ with the source node u and the sink node t. Define a set of labellings $\mathcal{L}_), \ldots, \mathcal{L}_\lceil$ such that each $L_i : E_i \to \mathbb{F}[\mathbf{x}]$. The labelling to every edge is thus, a polynomial in $\mathbb{F}[\mathbf{x}]$ of degree ≤ 1

- The vertices are partitioned into $d+1$ layers, i.e. $V = V_0 = s \sqcup V_1 \sqcup \ldots \sqcup V_d = t$ such that s adn t are the set of source and sink resp.
- Each edge e foes from V_{i-1} to V_i for some $i \in [d]$, so $E \subseteq \sqcup_{i \in [d]} V_{i-1} \times V_i$
- An edge e from V_{i-1} to V_i is labelled with an element $L_i = L\big|_{E_i}$
- The width of the ABP, denoted by w is $max_i |V_i|$
- The size of the ABP is the number of vertices $w^2 \cdot d$
- The polynomial computed by the ABP is $f = \sum_{p \in path(u,t} \prod_{e \in p} \mathcal{L}(e)$

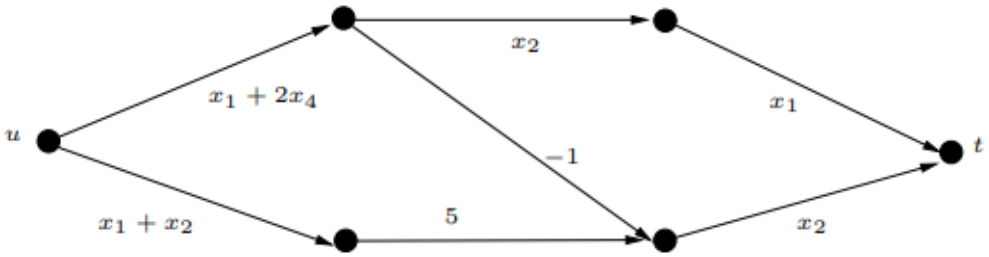

FIGURE 2.2: ABP computing $(x_1 + 2x_4)x_2x_2 - (x_1 + 2x_4)x_2 + 5x_2(x_1 + x_2)$

Computation :
The sum over all paths in a layered graph can be represented by an iterated matrix multiplication. Let w be the width of the ABP, $V = \sqcup V_i$ be the vertex set where $V_I = \{v_{i,j} | j \in [w]\}$

$$f = b^T(\prod_{i=1}^{q} A_i)c$$

$$\begin{aligned} b(l) &= wt(u, v_{0,l}) \ \ for \ \ 1 \leq l \leq w \\ A_i(k,l) &= wt(v_{i-1,k}, v_{i,l}) \ \ for \ \ 1 \leq l \leq w \ \ and \ \ 1 \leq i \leq w \\ T(k) &= wt(v_{d,k}, t) \ \ for 1 \leq k \end{aligned}$$

2.2.3 Read once Oblivious ABP

An ABP is called a *read-once oblivious ABP (ROABP)* if the edge weights in the different layers are univariate polynomials in distinct variables. Formally, the entries in D_i come from $\mathbb{F}[\mathbf{x}_{\pi(i)}]$ for all $i \in [d]$ where π is a permutation in the set $[d]$.

2.2.4 Set multi-linear circuit

A set multi-linear circuit is of the form

$$C(\mathbf{x}) = \sum_{i=1}^{k} \prod_{j=1}^{q} l_{i,j}(x_j)$$

where each $\mathbf{x}_1, \ldots, \mathbf{x}_q$ are disjoint set of variables and $l_{i,j}(x_j)$ is a linear polynomial in the variables $\mathbf{x}_j$ for each i, j. If we define vectors $v_{j,n} \in \mathbb{F}^k$ as $\mathbf{a}_{j,n} = (a_{1,j,n}, a_(w, j, n), \ldots, a_{k,j,n})$ then one can view the polynomial $f(x)$ as a dot product $(1, \ldots, 1) \cdot M(\mathbf{x})$ where

$$M(\mathbf{x}) = \prod_{j=1}^{q} (\mathbf{a}_{j,0} + \mathbf{a}_{j,1} x_{j,1} + \ldots + \mathbf{a}_{j,n} x_{j,n})$$

2.3 History of PIT Results

1. The first non-trivial deterministic test was found by [Ben-Or and Tiwari, 1988] which was a blackbox PIT for polynomials computed by depth-2 $(\Sigma \prod)$ circuits
2. In the context of ROABPs, [Raz and Shpilka, 2005] produced a poly(n, d, w) white-box algorithm for n variate, d- degree polynomials computed by a width-w ROABP with individual degree bounded by d.
3. [Saxena, 2008] showed that the log-variate ROABP model captures the depth 3 powering circuits by reducing the diagonal circuits to a sum of product of univariates.
4. [Agrawal et al., 2015] produced a quasi polynomial time $(ndw)^{O(loglogw)}$ time time sitting set for commutative ROABPs
5. [Agrawal et al., 2015] also gave a $O(ndw)^{logn}$ time hitting set for general ROABPs
6. The best blackbox test for diagonal circuits has time complexity $n^{O(loglogk)}$, which was shown by Forbes, Saptharishi, and Shpilka, 2014

Model	Time	Reference
$\sum\bigwedge\sum$	$(nd)^{O(\log n)}$	Agrawal *et al.* (2013)
—	$\text{poly}(d, 2^n)$	Forbes *et al.* (2018)
ROABP	$(ndr)^{O(\log n)}$	Forbes & Shpilka (2013b)
—	$n^{O(d \log r \log n)}$	Forbes *et al.* (2014)
—	$(ndr)^{O(\log n)}$	Agrawal *et al.* (2015)
—	$O(ndr^{\log n})$	Gurjar *et al.* (2017a)
Sum of c —	$(ndr)^{O(c2^c \log(ndr))}$	Gurjar *et al.* (2017b)
Border version	$(2^n(nd)^{\log n} r^{3^c \log n})^{O(c)}$	**This** work
—	$\text{poly}(d^c, r^{nc3^c})$	**This** work

FIGURE 2.3: Time complexities of different ROABP(n, d, w) models

Chapter 3

Structural Results

3.1 Background on ROABP

Recall, that a (layered) ROABP(n, d, w) is a computational model that uses exactly n matrices, one for each variable and the entries in the matrices are univariate polynomials in $\mathbb{F}[x_i]$ ROABPs can compute any monomial, and are closed under the summation operator.

Thus, every n variate,d-degree polynomial trivially has an ROABP of size $d^{O(n)}$ that can compute that ppolynoimial.

In general,there is a natural mapping between the order in which the ROABP reads the variables and the order of the matrices

Nisan's characterization **[NIS91]** furnished a critical observation realted to the order of variables In particular,

$$f = (x_1 + y_1)(x_2 + y_2)\dots(x_n + y_n)$$

- is coputatble by an ROABP whose width is polynomial in input size when the variable order is $(x_1, y_1, \dots, x_n, y_n)$
- the same polynomial requires an ROABP of width exponential in input size $(2^{\Omega(n)})$ when the variable order is changed to $(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)$

There, hence is a classification of the class of polynomials that are coputed by an ROABP

An ROABP with order permutation $\pi(n) = (x_1, \dots, x_n)$ can be expressed as $C = \mathbf{b}^T \cdot (\prod_{i\in[n]} M_i(x_i)) \cdot$ **c** The polynomial f, computed by this RO can thus be expressed as

$$f(\mathbf{x}) = \mathbf{b}^T \cdot (\prod_{i\in[n]} (A_{i,0} + A_{i,1}x_i + A_{i,2}x_i^2 + \dots + A_{i,d}x_i^d)) \cdot \mathbf{c} \tag{3.1}$$

3.1.1 Classifications

Defintion : ROABP[$\forall$] (n, d, w)
An n variate, d degree polynomial $f(\mathbf{x})$ such that $indv - deg(f) = d$ is said to have an RO of width w in every order, if there exists a width w ROABP that computes f for all permutations $\pi \in S(n)$ The class ROABP[$\forall$]$(n, d, w) = \cup$ ROABP[π](n, d, w) thus consists of all the n variate, d degree polynomials that are computable by a width w ROABP that reads input variables in *any* order

Defintion : ROABP[$\exists$] (n, d, w)
An n variate, d degree polynomial $f(\mathbf{x})$ such that $indv - deg(f) = d$ is said to have an RO of width w in the order $\pi(n)$, if there exists a width w ROABP that computes f in the order π. The class

ROABP$[\exists](n, d, w)$ consists of all the n variate, d degree polynomials that are computable by a width w ROABP that reads input variables in *some specific* π order

Defintion : commRO(n, d, w)
consists of all the n variate, d degree polynomials that are computable by a widthb w ROABP whose coefficient matrices $A_{i,j}$ commute with each other

Definition : diagRO(n, d, w)
consists of all the n variate, d degree polynomials that are computable by a widthb w ROABP whose all the $n(d+1)$ coefficient matrices $A_{i,j}$ are diagonal matrices.

3.1.2 Rank related results

Tensor rank :
Given a tensor $T : [d] \to \mathbb{F}$ of order n, it can be naturally expressed as a polynomial $f_T = \sum_{i \in [d]^n} T(i_, \ldots, i_n) x_q^{i_1}, \ldots, x_n^{i_n}$
For an n-variate, d-degree polynomial $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$, the Tensor rank of f, denoted by $TR(f)$ is the smallest width of a diagonal ROABP that computes it.

Waring Rank :
For an n-variate, d-degree polynomial $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$, the Waring rank of f, denoted by $WR(f)$ is defined as the size of the smallest depth 3 powering circuit that computes it.

Dimension of Partial Derivative space :
For an n-variate, d-degree polynomial $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$, its dimension of partial derivatives, denoted by $DPD(f)$, is the dimension of span of the Hasse derivatives $\partial_{\mathbf{e}} f \mid \mathbf{e} \in \mathbb{W}^n$.

3.2 Separation between ROABP subclasses

3.2.1 Nissan's characterization for ROABP by variable ordering

Nisan, 1991 **Lemma :** Let $f \in \mathbb{F}[\mathbf{x}, \mathbf{y}]$ such that

$$f(\mathbf{x}, \mathbf{y}) = \prod_i (\mathbf{x}_i + \mathbf{y}_i)$$

$$\text{Then, } dim(span\{coeff_f(\mathbf{x}^{\mathbf{a}})\}) = dim(span\{coeff_{\mathbf{y}}^{\mathbf{e}}\}) = 2^n$$

Proof :
clearly $\mathbf{a}, \mathbf{e} \in \{0,1\}^n$

$$coeff_f(y^{\mathbf{e}}) = \partial_{\mathbf{y}^{\mathbf{e}}}\big|_{\mathbf{y}=\vec{0}}$$
$$\mathbf{x}^{\vec{1}-\vec{\mathbf{e}}}$$

Thus, every coefficient of $\mathbf{y}^{\mathbf{e}}$ produces a distinct monomial in $\mathbb{F}[\mathbf{x}]$ and hence, $span(\partial_{\mathbf{y}^{\mathbf{e}}}(f) = span(\mathbf{x}^{1-\mathbf{e}} | \mathbf{e} \in \{0,1\}^n)) \geq 2^n$.

A symmetric argument holds for $dim(span(coeff_f \mathbf{x}^{\mathbf{a}}))$

But, as f is a multi-linear polynomial, the coefficient space must have size at most 2^n.
Hence, $dim(span\{coeff_f(\mathbf{x^a})\}) = dim(span\{coeff^{\mathbf{e}}_{\mathbf{y}}\}) = 2^n$

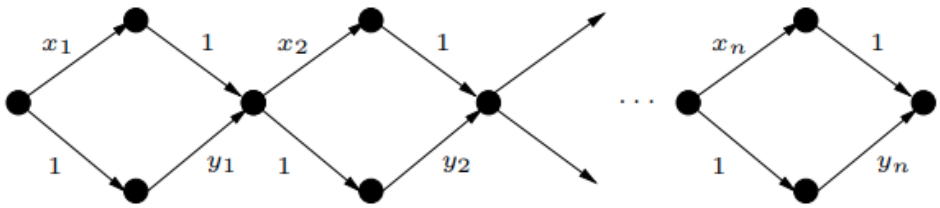

FIGURE 3.1: Variable ordering in ABP [Gurjar et al., 2017]

Furthermore, we can make the following inferences for $f(\mathbf{x}, \mathbf{y}) = \prod_i (\mathbf{x}_i + \mathbf{y}_i)$:

- For all permutations $\pi \in S(2n)$ of the variables $\mathbf{x}, \mathbf{y}$, f can be computed by an Roabp of width $\leq 2^n$
- There exists a permutation π of $\mathbf{x}, \mathbf{y}$ (characterized by the ordering of the variables in the product terms), such that any ROABP computing f in the variable order π has width 2. **Figure**
- There exists permutations π such that $\pi(x_i) < \pi(y_j) \forall i, j \in [n]$, then any ROABP computing f must have width $\geq 2^n$

3.2.2 Duality Trick

Saxena, 2008 Let f be a polynomial expressed in the form $f(\mathbf{x}) = (c_0 + \sum_{b_j=1}^{n} b_j x_j)^d$, then it can be written as a sum of product of univariates

$$\sum_{i=1}^{t} f_{i,1}(x_1) \cdot f_{i,2}(x_2) \dots f_{i,n}(x_n)$$

where $t = O(nd^2)$ and $F_{i,j}$ is a univariate polynomial in x_i for each i, j.

Proof :
Consider the polynomial $g(t; \mathbf{a}, \mathbf{x}) = \left((t + a_0)(t + a_1 x_1) \dots (t + a_n x_n) - t^{n+1} \right)^d$. Then $deg_t(g) = nd$ and $coeff_g(t^{nd}) = a_0 + \sum_{j=1}^{n} a_j x_j$ To extract the this coefficient we use the technique of polynomial interpolation.
Univariate interpolation : For a univariate n-variate, d-degree polynomial, any of its coefficient can be written as a linear combination of its $poly(n, d)$ number of evaluations. For the polynomial

g defined above, there exists constants $\{\alpha_i\}_{i=1}^{nd+1}$ and evaluation points $\{\beta_i\}_{i=1}^{nd+1}$ such that

$$\begin{aligned} coeff_g(t^i) &= \sum_{i=1}^{nd+1} \alpha_i g(\beta_i) \\ &= \sum_{r=0}^{d} \binom{d}{r} (-1)^{d-r} \beta_i^{(n+1)(d-r)} \cdot (\beta_i + a_0)^r \cdot (\beta_i + a_0 x_1)^r \cdots (\beta_i + a_n x_n)^r \end{aligned}$$

3.3 Investigation of ROABP subclasses

By scrutinizing the matrix structure [Equation 3.1] of the ROABP subclasses , we present some standard heirarchical results :

1. As the coefficient matrices of diagonal ROs are diagonal matrices, then they can capture any polynomial which is computed as *sum of product of univariates*
2. Saxena, 2008 showed that diagonal ROABPs can efficiently simulate $\Sigma \bigwedge \Sigma$
3. Nisan and Wigderson, 1996 showed an exponential separation between these classes for the polynomial witness $f(\mathbf{x}) = x_1 \cdot x_2 \cdots x_n$; which can be efficiently computed by a diagRO but requires exponential size $\Sigma \bigwedge \Sigma.=$, further implying that $\Sigma \bigwedge \Sigma \subsetneq diagRO$
4. From Nissan's characterisation by variable ordering **Sec 3.2.1** its is straightforward thaat $ROABP[\forall] \subsetneq ROABP[\exists]$
5. As all diagonall matrices commute wiith ech other, we have $diagRO(n, d, w) \subseteq commRO(n, d, w)$
6. Since the coefficient matrices in the $commRO$ are commutative, one can multiply the matrices in any order to compute the polynomial. Hence, $commRO$ can efficiently be subsumed by $ROABP[\forall](n, d, w)$

3.4 Heirarchy in ROABP models

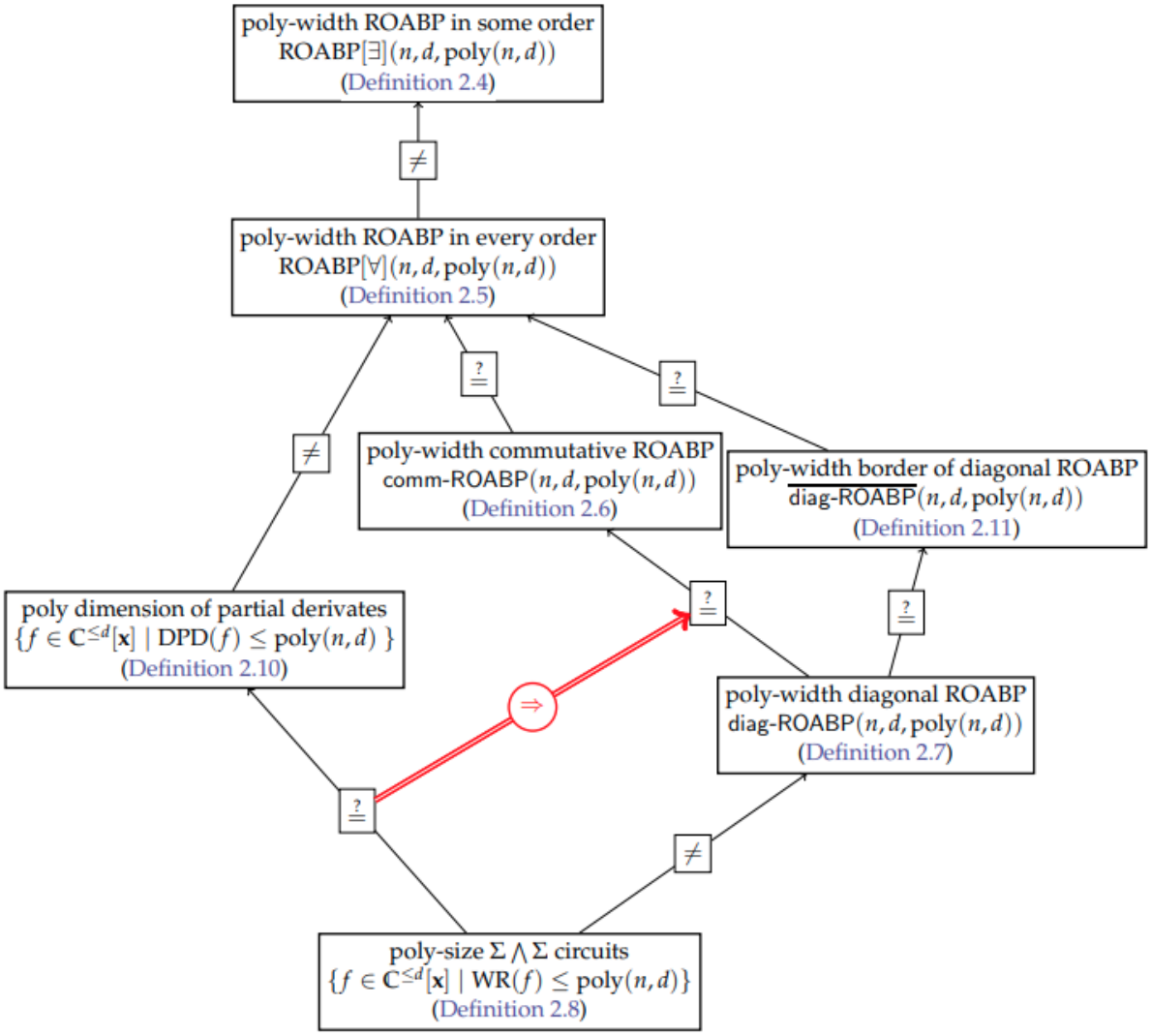

FIGURE 3.2: commRO, diagRO and $\Sigma \wedge \Sigma$ heirarchy [Ramya and Tengse, 2022]

Chapter 4

Separation between commRO & diagRO

4.1 Results

Structural relationship between commRO & diagRO

Theorem 4.1. Let f be an n-variate polynomial with the dimension of partial derivative space $\leq m$. Further, let $S(n,m)$ denote the smallest $\Sigma \wedge \Sigma$ - size required such an f. Then for all $n, d, z \in N$, commRO $(n,d,z) \subseteq$ diagRO$(n,d,S(z^2,z^2)nz^4)$

Consequences 1:
If the smallest $\Sigma \wedge \Sigma$ size required to compute any n- variate, d-degree polynomial f with $DPD(f) \leq m$ has size at most $(nds)^c$, for some constant $c > 0$ then the above theorem implies $commRO \subseteq diagRO$

Consequence 2:
If there exists an explicit polynomial that witnesses a super-polynomial seperation between the classes of commRO and diagRO; then that same polynomial will be a witness to super-polynomial separation between the dimension of partial derivative space and Waring Rank.

4.2 Proof Outline

1. Given a commRO(n, d, w) over the underlying field $\mathbb{F}$ and the variable set $\{x_1, \cdots, x_n\}$. Suppose this RO is of the structure $F(\mathbf{e}) = \mathbf{b}^T \cdot (\prod_{i\in[n]} (A_{i,0} + A_{i,1}x_i + A_{i,2}x_i^2 + \ldots + A_{i,d}x_i^d)) \cdot \mathbf{c}$
2. Let f be the polynomial that this ROABP computes. Then $f(x)$ is a linear combination (given by the entries in $\mathbf{b}\mathbf{c}^T$) of the entries of $F(x)$
3. Identify the set of coefficient matrices $A_1, \ldots, A_r$ that generarte the coefficient matrix ring. Since the dimension of the matrix algebra is at most w^2, we have $r \leq w^2, n(d+1)\ |\mathbb{F}[A_1, \ldots, A_r] = \mathbb{F}[A_{1,0}, \ldots, A_{1,d}, \ldots, A_{n,d}]$.
4. The ideal of dependencies of this generator set $\{A_1, \ldots, A_r\}$, its variety and the corresponding normal set are then used to identify the matrices in this algebra. Note that this ideal will have a finite variety, that is composed of the common zeros of the characteristic polynomials of all the generating matrices.
5. Any matrix $A_{i,j}$ can be expressed as a polynomial in $\{A_1, \ldots, A_r\}$ such that $\widetilde{A_{i,j}} = A_{i,j} \mod J$, where J is the prescribed ideal of dependecies of $\{A_1, \ldots, A_r\}$

6. As $\mathbb{F}[A_1, \ldots, A_r] \cong \frac{\mathbb{F}[\mathbf{t}]}{J}$, for any matrix $A_{i,j}$ we have $\widetilde{A_{i,j}}$ as a polynomial with elements in the normal set of J.
7. Let $H_{i,j}$ be the corresponding polynomial such that $A_{i,j} = H_{i,j}(A_1, \ldots, A_r)$, then $F(\mathbf{x})$ can be expressed a linear combination of $\mathbf{t}$ elements of $H(\mathbf{t}, \mathbf{x}) = \prod_{i \in [n]} \sum_{j \in [d} H_{i,j}(\mathbf{t}, x_i)$
8. Using results from **[Möller and Stetter, 1995, Marinari, Moeller, and Mora, 1993]**, Let Ω a derivative operator space corresponding to a zero-dimensional ideal J (i.e. an ideal with a finite variety). Then $\Omega = \bigcup \Omega_i$ where each Ω_i is spanned by a finite number of Derivative operators $\mathcal{D}_{i,j}$ characterized by the Normal set N_j
9. Then, any polynomial $H(\mathbf{t})$ its residue modulo this ideal, i.e. $\widetilde{h}(\mathbf{t}) = \widetilde{H}(\mathbf{t}) \mod J$ can be written as an m-linear combination of evaluations of its derivatives at the elements of the variety. $\widetilde{H}(\mathbf{t}) = \sum_{u \in |Var(j)|} \lambda_{u,*} \cdot \mathcal{D}_{u,*}(H)\big|_{\vec{\theta_u}}$
10. Using the methods of interpolation - univariate interpolation extended to term-wise homogeneous interpolation for multivariate polynomials, we configure a way to express each of these evaluations $\mathcal{D}_{u,*}(H)\big|_{\vec{\theta_u}}$ as linear combination of evaluations of $\widetilde{H}(\mathbf{t}, \mathbf{x})$ and hence as a linear combination of $\widetilde{H}(\mathbf{t}, \mathbf{x})$.
11. We employ the Waring decomposition of polynomial $\widetilde{H}(\mathbf{t}, \mathbf{x})$ to express its derivative evaluations at $\mathbf{0}$.
12. The number of these evaluation points of $\widetilde{H}$ required to compute the derivative evaluations turns out to be $poly(WR(H), deg(\mathcal{D}_{\widetilde{H}}))$.
13. To evaluate the polynomial derivatives at the points of points in the variety $\vec{\theta_i}$, appropriate shifts to the derivative evaluations at $\mathbf{0}$ can be done efficiently.
14. Using the hypothesis, that for an n variate polynomial the $WR(f) = poly(r, DPD(f))$, we have that $\mathcal{D}_{u,*}(H)\big|_{\vec{\theta_u}}$ can be expressed as $poly(r, DPD(H_{u,*}), deg(\mathcal{D}_{H_{u,*}}))$ evaluations
15. Since the space of Hasse derivatives is down closed under shifting, we have for each i, $|\Omega_i| = m_i$ such that $\bigcup \Omega_i \cong \frac{\mathbb{F}[\mathbf{t}]}{J}$. Thus $\sum m_i = dim(\frac{\mathbb{F}[\mathbf{t}]}{J})$, hence each $deg(\mathcal{D}_{H_{u,*}}) = m_i \leq dim(\frac{\mathbb{F}[\mathbf{t}]}{J})$.
16. Hence,for each $\mathcal{D}_{i,*} \in \Omega_i$, $\mathcal{D}_{i,*}\big|_{\theta_u}$ can be written as linear combination of $poly(r, m)$ evaluations of $H(\mathbf{t}, \mathbf{x})$
17. Given the hypothesis that the Waring rank of a polynomial is captured by the dimension of its partial derivatives upto polynomial factors, then $f(x)$ be written as $poly(n, d, w)$ evaluations of $H(\mathbf{t}, \mathbf{x})$.

Chapter 5

Proof

5.1 Algebraic structure of commRO

Algebraic Structure of commRO

Theorem 5.1. Suppose $f(x) = b^T(\prod_{i\in[n]}(A_{i,0} + A_{i,1}x_i + A_{i,2}x_i^2 + \ldots + A_{i,d}x_i^d))c$, such that f is computable by a commRO of width w. Then $f(x)$ can be expressed as linear combinations of the t- coefficients of an $(r+n)$-variate formal polynomial $G \in \mathcal{C}[t,x]$.

Proof
Let $F(x)$ denote the wXw matrix with entries in $\mathcal{C}[\vec{x}]$, so that $f(\vec{x}) = \vec{b}^T F(x)\vec{c}$. Define $\mathcal{A}$ as the commutative ring generated by the coefficient matrices $A_{i,j}$. Clearly, this ring $\mathcal{A}$ is a vector space over $\mathcal{C}$ with dimensions $\leq \ min(w^2, n(d+1))$. Let $\{A_1, A_2, \ldots, A_r\}$ be the generators of the ring $\mathcal{A}$. Let J be the ideal of dependencies for the coefficient matrix space defined as

$$J = \{h(\vec{t}) \in \mathbf{C}[\tilde{\mathbf{t}} \mid h(A_1, A_2, \ldots, A_r) = 0\}$$

Denote by N_J, the normal set of J. As $\frac{\mathbf{C}[\tilde{\mathbf{t}}]}{J} \cong \mathbf{C}[A_1, A_2, \ldots, A_r]$, so $N_J = \{\mathbf{t}^{\mathbf{e_i}}, \ldots, \mathbf{t}^{\mathbf{e_m}}\}$ with $m \leq w^2$.

the elements of the normal set N_J will be used to represent polynomials in the ring $\mathbb{C}[\mathcal{A}]$, which is isomorphic to the quotient ring of J with respect to $\mathbb{C}[$. In particular, $g \mod J$ can be written as a linear combination of monomials in N_J.

Define $H_{i,j}(\mathcal{A}) = A_{i,j}$ where $H_{i,j} \in \frac{\mathbf{C}[\mathbf{t}]}{J}$. Each layer of the commRO can then be represented as

$$H_i(\mathbf{t}, x_i) = \sum_j H_{i,j}(\mathbf{t}, x_i^j)$$

for $i \in [n]$ Note that, $deg_{\mathbf{t}}(H_i) = max\{deg_{\mathbf{t}}(H_{i,j})\}) \ \leq w^2$ for all i

Finally define $H(\mathbf{t}, \mathbf{x}) = \prod_i H_i(\mathbf{t}, x_i)$ and $\widetilde{H}(\mathbf{t}, \mathbf{x}) := H \mod J$

$$\widetilde{H} = \sum_{\mathbf{t^e} \in N_J} h_{\mathbf{e}}(\mathbf{x})\mathbf{t^e})$$

. These $h_{\mathbf{e}}$ are the $\mathbf{t}$- coefficients of G.

$$F(x) = H(\mathbf{A}, \mathbf{x}) = \widetilde{H}(\mathbf{A}, \mathbf{x}) = \sum_{\mathbf{e}\in N_J} \widetilde{h}_{\mathbf{e}}(\mathbf{x})\mathbf{A^e} \tag{5.1}$$

$$f(x) = \sum_{k,l\cdot\in[w]} b_k c_l \cdot F(x)[k,l] \tag{5.2}$$

$$= \sum_{k,l\in[w]} b_k c_l \cdot \sum_{\mathbf{e}\in N_J} (\widetilde{h}_{\mathbf{e}}(\mathbf{x})\mathbf{A}^{\mathbf{e}}[k,l]) \tag{5.3}$$

$$= \sum_{\mathbf{e}\in N_J} \Big(\sum_{k,l\in[w]} b_k c_l \cdot \mathbf{A}^{\mathbf{e}}[k,l] \Big) \widetilde{h}_{\mathbf{e}}(\mathbf{x}) \qquad = \sum_{\mathbf{e}\in N_J} \mu_{\mathbf{e}} \widetilde{h}_{\mathbf{e}}(\mathbf{x}) \tag{5.4}$$

5.2 t- coefficients as linear combination of derivatives

5.2.1 Leading Monomial ideal

For a polynomial $h(\mathbf{t})$, a monomial in the support of h is said to be the leading monomial of h (denoted by $LM(h)$), if for all $\mathbf{t}^{e'}$ we have $\mathbf{t}^{\mathbf{e}'} \prec \mathbf{t}^{\mathbf{e}}$ Further, we define the leading monomial for an ideal J, denoted by $LM(J) := \{LM(h) | h \in J\}$

5.2.2 Polynomial residues

Define $\widetilde{h}(\mathbf{t}) := h(\mathbf{t}) \mod J$
Observe that if $LM(\widetilde{h}) \notin \langle LM(J)\rangle$, then $supp(\widetilde{h}) \cap \langle LM(J)\rangle = \phi$
We, thereby decompose every polynomial modulo J. In other words, for any polynomial $h(\mathbf{t}) \in \mathbb{C}[t]$ and an ideal $J \in \mathbb{C}[t]$, we have

$$h(\mathbf{t}) = h_J(\mathbf{t}) + \widetilde{h}(\mathbf{t})$$

, such that $LM(\widetilde{h})$ is not contained in the ideal $\langle LM(J)\rangle$.

Corresponding to an ideal $J \subsetneq \mathbb{C}[t_1, t_2, \ldots, t_r]$, the **Normal set** of J is defined as $N_J := \{\mathbf{e} \mid \mathbf{e} \in \mathbb{N}^r, \mathbf{t}^{\mathbf{e}} \notin \langle LM(J)\rangle\}$. Residue of any polynomial $h(\mathbf{t}) \mod J$ can be written as a linear combination of monomials in N_J, further $|N_J| = dim(\frac{\mathbb{F}[\mathbf{t}]}{J})$

5.2.3 Normal Set and Characterizing Derivative Operator Spaces N_J

[Marinari, Moeller, and Mora, 1993]

Given $J \subseteq \mathbb{C}[\mathbf{t}]$ an ideal with variety $\mathbf{V}(J) = \{\vec{\theta_0}, \vec{\theta_1}, \ldots, \vec{\theta_s}\}$ and corresponding Normal Set $N_J = \{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$. Let $\Omega_1, \ldots, \Omega_m$ be the characterising Hasse Derivative space with each Ω_i spanned by $\{\mathcal{D}_{i,1}, \ldots, \mathcal{D}_{i,b_i}\}$, such that $|N_J| = \sum_i b_i = m$

Then, we have a small finite set of constants, such that for any polynomial $h(\mathbf{t}) \in \mathbb{C}[\mathbf{t}]$ with $\widetilde{h} = h \mod J$ we have an explicit formulation of $coeff_{\mathbf{e}}(\widetilde{h})$ for all $\mathbf{e} \in N_J$. In particular, we have

$$coeff_{\mathbf{e}}(\widetilde{h}) = \sum_{i,v} \lambda^{(\mathbf{e})} \mathcal{D}_{i,v}(h) \Big|_{\theta_i}$$

This explicit set of m^2 constants $\lambda^{(e)}$ corresponding to the characterizing Hasse operator spaces is a consequence of a result by Moller and Stetter.

Multivariate roots & multiplicities

Theorem 5.2. Möller and Stetter, 1995 In particular, consider the ideal $J = \{h(\vec{t}) \in \mathbf{C}[\tilde{\mathbf{t}} \mid h(A_1, A_2, \ldots, A_r) = 0\}$, $var(J) = \{\theta_0, \ldots, \theta_s\}$ and the corresponding Normal set $N_J = \{\mathbf{e}_1, \ldots, \mathbf{e}_z \ \ |z \leq dim(\frac{F[\mathbf{t}]}{J})\}$. Corresponding to each $\theta_i \in var(J)$, we have the characterizing derivative operator space Ω_i spanned by $\{\mathcal{D}_{i,1}, \ldots, \mathcal{D}_{i,b_i}\}$, such that $|N_J| = \sum_i b_i = m = dim(\frac{F[\mathbf{t}]}{J})\}$. Then we have, a set of z^2 constants $\lambda_{i,j} | i, j \in [z]$.

Equivalently we have z polynomials ϕ_i with $DPD(\phi_i) \leq z$ with each ϕ_i characterizing the derivative operator $\sum_{j \in [z]} \lambda_{i,j} \mathcal{D}_{i,j}$. Then $\widetilde{H} = H \mod J$ can then be expressed as a linear combination of evaluations of these polynomials at the elements of $var(J)$.

$$coeff_{\mathbf{e}}(\widetilde{H}) = \sum_{i,v} \lambda^{(\mathbf{e})} \mathcal{D}_{\phi_{i,v}}(h) \Big|_{\theta_i}$$

5.3 Evaluations of Derivatives

5.3.1 Generic polynomials

First let us show the idea through generic polynomials f, g that satisfy certain conditions and then we will use a similar technique on the polynomial $H(\mathbf{t}, \mathbf{x})$ under consideration. Let $g, h \in \mathbb{C}[\mathbf{t}]$ be polynomials of degree at most d ans further suppose $WR(g) \leq \tau$. Then we show that $\mathcal{D}_g(h)\big|_{\mathbf{0}} = \mathcal{D}_h(g)\big|_{\mathbf{0}}$ can be expressed as a linear combination of $h(\xi_1), \ldots, h(\xi_q)$ where $Q = \{\xi_q\}, |Q| = O(w \cdot d)$

We express both h and g as a linear combination of their homogeneous components. As the Hasse derivative space respects linearity, we use these homogeneous components to characterize $\mathcal{D}_h(g)$ and $\mathcal{D}_g(h)$.

$$\begin{aligned} g = \sum_{0 \leq i' \leq d} g_{i'} \ , \ h = \sum_{0 \leq i \leq d} h_i \\ \mathcal{D}_g(h)\big|_{\mathbf{0}} = \sum_{0 \leq i' \leq d} \sum_{0 \leq i \leq d} \mathcal{D}_{g_{i'}}(h_i)\big|_{\mathbf{0}} \\ \mathcal{D}_{g_{i'}}(h_i) = 0 \ \ for(i < i') \\ \mathcal{D}_{g_{i'}}(h_i)\big|_{\mathbf{0}} = 0 \ \ for(i > i') \\ \mathcal{D}_g(h)\big|_{\mathbf{0}} = \sum_{0 \leq i \leq d} \mathcal{D}_{g_i}(h_i)\big|_{\mathbf{0}} = \sum_{0 \leq i \leq d} \mathcal{D}_{h_i}(g_i)\big|_{\mathbf{0}} \end{aligned} \tag{5.5}$$

To show that $O(WR(g) \cdot max\{deg(g), \ deg(h)\})$ evaluation points are sufficient to express h; we look at the Waring decomposition of h .**[Saxena, 2008]**.

$$h = \sum_{j \in [WR(h)]} (c_j + \langle \mathbf{b_j}, \mathbf{t} \rangle)^{d_j}$$

$$
\begin{aligned}
h &= \sum_{j\in[\tau]} \binom{d_j}{k} c_j^{d_j-k} \cdot (\langle \mathbf{b_j}, \mathbf{t}\rangle^k) \\
\mathcal{D}_{h_i}(g_i)\big|_{(0)} &= \sum_{j\in[\tau]} \binom{d_j}{k} c_j^{d_j-k} \cdot \mathcal{D}_{g_i}(\langle \mathbf{b_j}, \mathbf{t}\rangle^k)\big|_{\mathbf{0}} \\
&= \sum_{j\in[\tau]} \binom{d_j}{k} c_j^{d_j-k} \cdot \sum_{\mathbf{e}\in supp(g)} coeff_{\mathbf{e}}(h_i) \cdot \partial_{\mathbf{e}}(\langle \mathbf{b_j}, \mathbf{t}\rangle^k)\big|_{\mathbf{0}} \\
&= \sum_{j\in[\tau]} \binom{d_j}{k} c_j^{d_j-k} \cdot \sum_{\mathbf{e}\in supp(g)} coeff_{\mathbf{e}}(h_i) \cdot i! \cdot \mathbf{b}_j^{\mathbf{e}} \qquad = \sum_{j\in[\tau]} \xi_{i,j} \cdot h_i(\mathbf{e_j})
\end{aligned}
$$

$$
\mathcal{D}_g(h)\big|_{\mathbf{0}} = \mathcal{D}_h(g)\big|_{\mathbf{0}} = \sum_{0\le i\le d} \sum_{j\in[\tau]} \xi_{i,j} \cdot h_i(\mathbf{b_j}) \tag{5.6}
$$

To obtain $\mathcal{D}_g(h)\big|_{\mathbf{0}}$ as linear combination of evaluations of g, we use the method of interpolation of homogeneous components of a polynomial.

$$
\begin{aligned}
h_i(\mathbf{b}_j) &= coeff_{v^i}\big(h(v\cdot b_{j,1}\ ,\ \dots\ ,\ v\cdot b_{j,n})\big) \\
h_i(\mathbf{b}_j) &= \sum_{0\le k\le d} \lambda_{i,k} h(\mathbf{v}\cdot\mathbf{b}_j) \qquad \text{[univariate interpolation]} \\
\mathcal{D}_{h_i}(g_i)\big|_{(0)} &= \sum_{j\in[w]} \Big(\sum_{0\le k\le d} \lambda_k \cdot h(v_k\cdot b_{j,1}\ ,\ \dots\ ,\ v_k\cdot b_{j,n}) \Big) = \sum_{j\in[w]} \Big(\sum_{0\le k\le d} \lambda_k \cdot h(\mathbf{v}_k\cdot\mathbf{b}_j) \Big) \\
\mathcal{D}_g(h)\big|_{\mathbf{0}} &= \sum_{0\le i\le d} \sum_{j\in[w]} \xi_{i,j} \cdot \Big(\sum_{0\le k\le d} \lambda_{i,k} \cdot h(\mathbf{v}_k\cdot\mathbf{b}_j) \Big) \\
\mathcal{D}_g(h)\big|_{\mathbf{0}} &= \sum_{j\in[w]} \sum_{0\le k\le d} \Big(\sum_{0\le j\le d} \xi_{i,j}\cdot\lambda_{i,k} \Big) \cdot h(\mathbf{v}_k\cdot\mathbf{b}_j) \\
\mathcal{D}_g(h)\big|_{\mathbf{0}} &= \sum_{j\in[w]} \sum_{0\le k\le d} \gamma_{j,k} \cdot h(\mathbf{v}_k\cdot\mathbf{b}_j) \\
\mathcal{D}_g(h)\big|_{\mathbf{0}} &= \sum_{\eta\in[w\cdot d]} \gamma_e \cdot h(\delta_\eta)
\end{aligned}
$$

for $[\delta_1, \dots, \delta_{w\cdot d} \in \mathbb{F}^n]$ and $w\cdot d = WR(h)\cdot max\{deg(g), deg(h))\}$

5.3.2 For the polynomial $H(\mathcal{A}, \mathbf{x}) = F(\mathbf{x})$ corresponding to the space of coefficient matrices

Recall that $H(\mathbf{t}, \mathbf{x}) = \prod_{i\in[n]} \sum_{j\in[d} H_{i,j}(\mathbf{t}, x_i)$ for the underlying coefficient ring $\mathcal{A}$ and the zero -dimensional ideal of dependencies J, $H_{i,j}(\mathcal{A} = A_{i,j})$
Consider the polynomials ϕ_i obtained after expressing H as a linear combination of evaluations of the derivative operator spaces characterized by the $var(J)$ [given by Theorem 2.2 : Multivariate roots & multiplicities]

Sufficient number of Evaluation points

Theorem 5.3. Given that $coeff_{\mathbf{e}}(\widetilde{H}) = \sum_{i,v} \lambda^{(\mathbf{e})} \mathcal{D}_{\phi_{i,v}}(h)\Big|_{\theta_i}$. Let $\tau = WR(\phi) = max\{WR(\phi_{i,v})\}$ And let $d = max\{deg(H), deg(\phi_{i,v})\}$, then there exists at most $r = \tau \cdot d$ points $\delta \in \mathbb{F}^n$ such that

$$\mathcal{D}_{\phi_{i,v}}(H)\big|_{\theta_i} = \sum_{\eta \in [w \cdot d]} \gamma_e \cdot H(\delta_\eta)$$

This, we iterate on $\phi_{i,v}$ for all $i \in |var(J)|$

1. According to our hypothesis, for r-variate polynomials $\phi_{i,v}$ with $DPD(\phi_{i,v}) \leq m$, the $WR(\phi_{i,v}) = S(r,z)$.
2. Also, as $deg(H_i) \leq w^2$, the $deg(H) = deg(\prod_1^n H_i) \leq n \cdot w^2$

Thus, in total we require $O(\tau \cdot d) \equiv O(S(r,z) \cdot n \cdot w^2)$ evaluations to obtain $\mathcal{D}_{\phi_{i,v}}(H)\big|_{\theta_i}$

5.3.3 Consolidating all the results

$$\begin{aligned}
f(\mathbf{x}) &= \sum_{\mathbf{e} \in N_J} \mu_{\mathbf{e}} \widetilde{h_{\mathbf{e}}}(\mathbf{x}) \\
&= \sum_{\mathbf{e} \in N_J} \mu_{\mathbf{e}} \Bigg(\sum_{\substack{i \in var(J) \\ v \in [\Omega_i]}} \lambda^{\mathbf{e}}_{i,v} \cdot \mathcal{D}_{i,v}(H)\big|_{\theta_i} \Bigg) \\
&= \sum_{\substack{i \in var(J) \\ v \in [\Omega_i]}} \mu'_{i,v} \cdot \Big(\mathcal{D}_{i,v}(H)\big|_{\theta_i} \Big) && \text{Roots + Multiplicity} \\
&= \sum_{\substack{i \in var(J) \\ v \in [\Omega_i]}} \mu'_{i,v} \cdot \Bigg(\sum_{\eta=1}^{nw^2 \cdot S(r,z)} \gamma_v \cdot H(\delta_\eta) \Bigg) \\
&= \sum_{\eta=1}^{nw^2 \cdot S(r,z)} \Big(H(\delta_\eta, \mathbf{x}) \Big) \cdot \gamma_\eta \cdot \sum_{\substack{i \in var(J) \\ v \in [\Omega_i]}} \mu'_{i,v}
\end{aligned}$$

$$f(\mathbf{x}) = \sum_{\eta'=1}^{nw^2 \cdot S(r,z)} \kappa'_\eta \cdot \prod_i^n \Big(H_i(\delta'_\eta, x_i) \Big) \tag{5.7}$$

5.4 Conclusion

Note that in **eqn. 2.7** , we have expressed $f(\mathbf{x})$ as a sum of product of diagonal matrices $H_i(\mathbf{t}, \mathbf{x})$, where the formal variable $\mathbf{t}$ takes input $\mathcal{A} = \{A_1, \ldots, A_r\}$, the ideal of dependencies of the coefficient matrix space $spanA_{i,j}$. The number of evaluations sufficient for complete characterization are polynomial in $m, w, S(r,z)$.

As $r, z \leq w^2$, we have that $S(r,z) \leq S(w^2, w^2)$.

Given the hypothesis in Theorem 4.1 (Consequence 1) **[DPD(f) $\leq_p$ WR(f)]**, we thus construct a diagonalROABP of size $O(w^2 \cdot S(w^2, w^2) \cdot nw^2)$ that simulates a commutative ROABP with width w, computing a polynomial an n- variate, d-degree polynomial f.

Chapter 6

Polytopes and PIT

Main result in this section revolves around the construction of an explicit weight assignment family that isolates log-variate polynomials that have a low dimension of partial derivatives. $\Sigma^{[s]} \bigcap^{[d]} \Sigma^{[n]}$ will denote an n variate depth 3 powering circuit with degree $\leq d$ and top fan in $\leq s$

6.1 Polytopes and Cone closed basis

F-cone of a monomial

Definition 6.1. Let $\mathbb{F}$ be the underlying Field and $\mathbf{m}$ be a monomial over the variable set $\{x_1, \ldots, x_r\}$.

$$\mathbb{F} - cone(\mathbf{m}) = \{\partial_{\mathbf{e}}\mathbf{m} | \mathbf{e} \in \mathbb{N}^r, \partial_{\mathbf{e}}\mathbf{m} \neq 0\} \tag{6.1}$$

6.1.1 Newton Polytopes

Let $S = \{u_1, \ldots, u_m\} \mid u_i \in \mathbb{R}^d$, then we define the convex span or convex hull of S as

$$CS(S) = (\ \Sigma_i \alpha_i u_i \mid \alpha_i \in \mathbb{R},\ \alpha_i \geq 0, \Sigma_i \alpha_i = 1)$$

A set $P \subset \mathbb{R}^d$ is called a convex set in $\mathbb{R}^d$ if for any two points $u, v \in P$ and any $0 \leq \alpha \leq 1$

$$\alpha u + (1 - \alpha)v \in P$$

A convex set (say, $P \subset \mathbb{R}^d$) is called a polytope is there exists a finite set $S \subset \mathbb{R}^d$ such that $P = CS(S)$

For any polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$, the Newton Polytope P_f is defined as

$$P_f = CS(supp(f))$$

$$P_f = CS(\{\mathbf{e} \in \mathbb{N}^n \mid coeff_f(\mathbf{x}^{\mathbf{e}}) \neq 0\})$$

6.1.2 Vertices of a polytope & Minima of Linear functions

Suppose $P \subseteq \mathbb{R}^n$ be a convex polytope. A point $v \in P$ is a vertex of P if it cannot be expressed as a non-trivial convex combination of other points in P.

$\mathbf{v} \in P$ is a vertex if there is now $\mathbf{u}, \mathbf{w} \in P \backslash \{\mathbf{v}\}$ suchhthat for any $0 \leq \alpha \leq 1$

$$v = \alpha u + (1 - \alpha)w$$

For any convex polytope $P \subseteq \mathbb{R}^n$, the following hold

- For any $\mathbf{u} \in \mathbb{R}^n$ and a point $\mathbf{e} \in P$, there exists a vertex v of P, such that $\mathbf{u}^T\mathbf{v} \leq \mathbf{u}^T\mathbf{e}$
- If $\mathbf{v}$ is a vertex of P, then there exists a vector $\mathbf{a} \in \mathbb{R}^n$ such that the linear function $L_{\mathbf{a}} : \mathbf{x} \rightarrow \mathbf{a}^T\mathbf{x}$ is uniquely minimized on p at $\mathbf{v}$
- If $\mathbf{v}$ is a vertex of P and if $\mathbf{a}^T\mathbf{v} < \mathbf{a}^T\mathbf{v}'$ for all vertices $\mathbf{v}' \neq \mathbf{v}$ of P, then the linear function $L_{\mathbf{a}} : \mathbf{x} \rightarrow \mathbf{a}^T\mathbf{x}$ is uniquely minimized on P at $\mathbf{v}$

6.2 Constructing IWA through Newton Polytopes

6.2.1 Basis Isolating Weight Assignment (BIWA)

A weight function $wt : x \rightarrow W$ is termed as a Basis isolating weight assignment for a polynomial $f(\mathbf{x}) \in F[\mathbf{x}]$, if there exists a set of monomials $S \subseteq supp(f)$ whose coefficients form a basis for the coefficient sapce of $f(\mathbf{x})$ such that the following hold

1. for any $e, e' \in S, wt(e) \neq wt(e')$
2. for any monomial $\mathbf{e} \in supp(f) \backslash S, coeff_f(\mathbf{e}) \in span\{coeff_f(\mathbf{e}') \mid \mathbf{e}' \in S, w(\mathbf{e}) < w(\mathbf{e}')\}$

A weight asignment that gives distinct weights to the all the monomials of a polynomial is necessarily a basis isolating weight assignment. However, in certain cases, it might inovlve exponentially large weights. So we try to use the structural properties of the circuit to construct and efficien BIWA

6.2.2 Hitting sets via Basis Isolation

6.2.3 Lemma

Let $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be an n-variate polynomial of degree d and let P_f be its Newton Polytope. Then for any vertex $\mathbf{e}$ of P_f we have $dim(\partial^*(f) \geq dim(\partial^*(\mathbf{x}^{\mathbf{e}})) = |\mathbb{F} - cone(\mathbf{e})|)$ As $\mathbf{e}$ is a vertex of P_f, there must exist a vector $\mathbf{w}$ such that the linear function $L_{\mathbf{w}}$ is uniquely minimized at $\mathbf{e}$.

Counting monomials with a low cone

Theorem 6.1. For a field $\mathbb{F}$ with characteristic 0 (or large characteristic)let $d, k \in \mathbb{N}$ be appropriately large enough. Then the number of n-variate, degree d monomials that have $\mathbb{F} - CS(\mathbf{e}) \leq k$ is at most $\binom{n}{logk}k^2$

Lemma : For $0 \leq k \leq n, k \in \mathbb{Z}$, we have $\sum_{i=0}^{k} \binom{n}{i} \leq \binom{en}{k}^k$

Proof

The number of n-variate monomials with cone-size $\leq k$ is $O(\eta k^2)$ where $\eta = (\frac{3n}{logk})^{logk}$

Let $N(k, l)$ denote the number of cone-size monomials with support X_l of size l. The exponent of each $x_i \in X_l$ is atleast ! and $\leq k - 1$, which gives the following recurrence

$$N(k, l) \leq \sum_{i=2}^{k} N(k/2, l - 1)$$

It can be proved by induction that $N(k,l) < k^2$ satisfies the above recurrence.

From the definition of cone, a cone-size $\leq k$ must have support size $\leq l = \lfloor logk \rfloor$. From here we can getm the number of possible support sets with cone-size $\leq k$ to be $\sum_{i=0}^{l} \binom{n}{i} =\leq (\frac{3n}{logk})^{logk}$

Isolating Weight Assignments

Theorem 6.2. Suppose $\mathbb{F}$ be a field of characteristic zero (or large characteristic). Let $d, k \in \mathbb{N}$ be appropriately large enough with $(C)(n,k,d)$ denoting the class of n-variate, d-degree polynomials over $\mathbb{F}$, such that $\partial * (f) \leq k, \forall f \in (C)$. Then there is an explicit family $W(k,d)$ which isolates $\mathcal{C}()k,d$ and consists of $poly(k, logd)$ weight assignments

Proof
Pick an arbitrary polynomial $f \in C(k,d)$ and let P_f be its Newton Polytope. $\mathbb{F} - CS(\mathbf{e})$ for the the vertices $\mathbf{e}$ of P_f will have size at most k. From previous results, we also have an upper bound on the number of such vertices. Using **Theorem 6.2**, we can conclude that P_f has at most $((\binom{n}{logk}) \cdot k^2)$.

Furthermore, for any polynomial f where $n = O(logk)$ in $C(n,d,k)$ the number of these vertices are $poly(k)$. As the choice of such an $f \in C$ was arbitrary, we have a family of weight assignments $W(n,d,k)$, where $n = O(log(k))$ which isolates C.

6.3 Conclusion

6.3.1 Hitting sets for low partials

As a corollary to the above theorem, we have for all large enough $k, d \in \mathbb{N}$ and $n = O(logk)$, the class $C(k,d)$ of n-variate, degree d polynomials over $\mathbb{F}$ with the dimension of partial derivative space at most k, has a hitting set of size $poly(k,d)$

6.3.2 Log variate Depth 3 Powering circuits

For large enough $s, d \in \mathbb{N}$ and $n = O(logk)$, the class $\Sigma^{[s]} \cap^{[d]} \Sigma$ circuits over $\mathbb{F}$ has a hitting set of size $poly(s,d)$.

Bibliography

Agrawal, Manindra (2005). "Proving lower bounds via pseudo-random generators". In: *International Conference on Foundations of Software Technology and Theoretical Computer Science*. Springer, pp. 92–105.

Agrawal, Manindra et al. (2015). "Hitting-sets for ROABP and sum of set-multilinear circuits". In: *SIAM Journal on Computing* 44.3, pp. 669–697.

Ben-Or, Michael and Prasoon Tiwari (1988). "A deterministic algorithm for sparse multivariate polynomial interpolation". In: *Proceedings of the twentieth annual ACM symposium on Theory of computing*, pp. 301–309.

Forbes, Michael A, Ramprasad Saptharishi, and Amir Shpilka (2014). "Hitting sets for multilinear read-once algebraic branching programs, in any order". In: *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pp. 867–875.

Gurjar, Rohit et al. (2017). "Deterministic identity testing for sum of read-once oblivious arithmetic branching programs". In: *computational complexity* 26.4, pp. 835–880.

Heintz, Joos and Claus-Peter Schnorr (1980). "Testing polynomials which are easy to compute". In: *Proceedings of the twelfth annual ACM Symposium on Theory of Computing*, pp. 262–272.

Marinari, Maria Grazia, H. Michael Moeller, and Teo Mora (1993). "Gröbner bases of ideals defined by functionals with an application to ideals of projective points". In: *Applicable Algebra in Engineering, Communication and Computing* 4.2, pp. 103–145.

Möller, H Michael and Hans J Stetter (1995). "Multivariate polynomial equations with multiple zeros solved by matrix eigenproblems". In: *Numerische Mathematik* 70.3, pp. 311–329.

Nisan, Noam (1991). "Lower bounds for non-commutative computation". In: *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, pp. 410–418.

Nisan, Noam and Avi Wigderson (1996). "Lower bounds on arithmetic circuits via partial derivatives". In: *Computational complexity* 6.3, pp. 217–234.

Ramya, C and Anamay Tengse (2022). "On Finer Separations Between Subclasses of Read-Once Oblivious ABPs". In: *arXiv preprint arXiv:2201.06432*.

Raz, Ran and Amir Shpilka (2005). "Deterministic polynomial identity testing in non-commutative models". In: *computational complexity* 14.1, pp. 1–19.

Saxena, Nitin (2008). "Diagonal circuit identity testing and lower bounds". In: *International Colloquium on Automata, Languages, and Programming*. Springer, pp. 60–71.

Zippel, Richard (1979). "Probabilistic algorithms for sparse polynomials". In: *International symposium on symbolic and algebraic manipulation*. Springer, pp. 216–226.