On Finer Separations Between Subclasses of Read-Once Oblivious ABPs

C. Ramya
Chennai Mathematical Institute, India

Anamay Tengse
Department of Computer Science, University of Haifa, Israel

Abstract

Read-once Oblivious Algebraic Branching Programs (ROABPs) compute polynomials as products of univariate polynomials that have matrices as coefficients. In an attempt to understand the landscape of algebraic complexity classes surrounding ROABPs, we study classes of ROABPs based on the algebraic structure of these coefficient matrices. We study connections between polynomials computed by these structured variants of ROABPs and other well-known classes of polynomials (such as depth-three powering circuits, tensor-rank and Waring rank of polynomials).

Our main result concerns *commutative ROABPs*, where *all* coefficient matrices commute with each other, and *diagonal ROABPs*, where all the coefficient matrices are just diagonal matrices. In particular, we show a somewhat surprising connection between these models and the model of *depth-three powering circuits* that is related to the *Waring rank* of polynomials. We show that if the *dimension of partial derivatives* captures *Waring rank* up to polynomial factors, then the model of *diagonal ROABPs* efficiently simulates the seemingly more expressive model of *commutative ROABPs*. Further, a *commutative ROABP* that cannot be efficiently simulated by a *diagonal ROABP* will give an explicit polynomial that gives a super-polynomial separation between *dimension of partial derivatives* and *Waring rank*.

Our proof of the above result builds on the results of Marinari, Möller and Mora (1993), and Möller and Stetter (1995), that characterise rings of commuting matrices in terms of polynomials that have small dimension of partial derivatives. The algebraic structure of the coefficient matrices of these ROABPs plays a crucial role in our proofs.

2012 ACM Subject Classification Theory of computation → Algebraic complexity theory

Keywords and phrases Algebraic Complexity Theory, Algebraic Branching Programs, Commutative Matrices

Digital Object Identifier 10.4230/LIPIcs.STACS.2022.53

Related Version *Full Version*: `https://anamay.bitbucket.io/assets/docs/papers/ROABP-Hierarchy.pdf`

Funding *C. Ramya*: Research supported by INSPIRE Faculty Fellowship of DST and by a grant from the Infosys Foundation. Part of this work was done when at the Tata Institute of Fundamental Research, Mumbai, India (DAE project 12-R&D-TFR-5.01-0500).
Anamay Tengse: Research supported by the Israel Science Foundation (grant No. 716/20). Part of this work was done when at the Tata Institute of Fundamental Research, Mumbai, India, as a student (DAE project no. 12-R&D-TFR-5.01-0500), and later as a visitor (Prof. Prahladh Harsha's Swarnajaynti fellowship and Prof. Arkadev Chattopadhyay's Microsoft Research funds).

Acknowledgements We thank Ramprasad Saptharishi for numerous insightful discussions about the various structured models, which motivated this work. We also thank Mrinal Kumar for his helpful comments about our work which helped us in enchancing the presentation.
We thank Manoj Gopalakrishan and the organisers of *Thursday Theory Lunch* at IIT Bombay for organising a talk by Debasattam Pal, where we first came across the work of Möller and Stetter (1995) that essentially led to the main results in this paper.
We thank the anonymous reviewers for their valuable inputs on the earlier version of the paper.

1 Introduction

The central question in *algebraic complexity theory*: the theory concerning computation of polynomials, is to understand the most efficient way of computing a polynomial $f(x_1, \ldots, x_n)$ using the basic arithmetic operations of addition and multiplication. One of the earliest works to study the computational complexity of an *explicit* polynomial is perhaps the famous work of Strassen [22] on matrix multiplication. However, the seminal work of Valiant [23] that proposed the "VP vs VNP" question (the algebraic analogue of P vs NP) is widely regarded as the starting point of algebraic complexity theory.

Algebraic circuits are a fundamental model for computing polynomials, and the complexity of a polynomial is determined by the *size* of the smallest circuit that computes it. This definition also coincides with the fewest number of arithmetic operations required to evaluate a polynomial. Valiant's above mentioned work however, uses the model of *algebraic branching programs (ABPs)* to capture *efficiently computable polynomials*. Informally, an ABP computes a polynomial $f(\mathbf{x})$ as the $(1,1)$th entry of a product of matrices, each of which has linear forms in the $\mathbf{x}$ variables as its entries. While VP is the class of n-variate polynomials having poly(n) size algebraic circuits, the class of n-variate polynomials that have an ABP of size poly(n) is called VBP. The class VBP is known to be a subclass of VP, and at the moment it is unclear if this inclusion is strict. The VBP vs VNP question remains a central question in algebraic complexity theory as it is captured by the "determinant vs permanent" question (see e.g. [10]).

Although proving strong lower bounds against algebraic circuits seems currently unattainable, even proving lower bounds against ABPs remains a challenging task. In fact, even a super-quadratic lower bound against ABPs will be a massive improvement over the state of the art ([1, 2]). A significant amount of work in the area has therefore focused on analysing more structured variants of ABPs which could potentially be easier to tackle. Indeed, a celebrated result of Nisan [14] gives an exact characterisation of the complexity of a *non-commutative ABP* computing any non-commutative polynomial[1]. This characterisation yields a $2^{\Omega(n)}$ lower bound against non-commutative ABPs for the determinant, which among other things, highlights the power of commutativity.

We now turn to the protagonists of our work, Read-once Oblivious ABPs (ROABPs), which are the commutative analogues of non-commutative ABPs. ROABPs were first introduced by Forbes and Shpilka [6], in the context of *polynomial identity testing*: another central problem in algebraic complexity, which we discuss in more detail in the full version. An ROABP is an algebraic branching program that uses exactly n matrices, one for each variable; and the entries in the matrix corresponding to an x_i are univariate polynomials from $\mathbb{C}[x_i]$ (formally defined in Definition 9). It is easy to check that ROABPs can compute any monomial, and are closed under taking sums. Thus, every n-variate, degree-d polynomial trivially has an ROABP of size $d^{O(n)}$. On the other hand, Nisan's characterisation [14] for non-commutative ABPs also extends to ROABPs, and hence most of the strong lower bounds against non-commutative ABPs can be suitably translated to ROABPs.

Since all ROABPs use n matrices, the parameter of interest is the *width* of an ROABP, which is the maximum dimension of any of the underlying matrices. Furthermore, since every matrix in an ROABP is associated with exactly one variable in $\{x_1, \ldots, x_n\}$, one can naturally identify an order $\sigma \in S_n$ (permutation on $\{x_1, \ldots, x_n\}$) in which the ROABP "reads the variables". Indeed, there are polynomials which are computable by poly(n)-width ROABPs

[1] A non-commutative polynomial is one in which the variables do not commute, i.e. $xy \neq yx$.

in one order, but require exponential width in a different order. In fact, a straight-forward application of Nisan's characterisation shows that the $2n$-variate polynomial $(x_1 + y_1)(x_2 + y_2) \cdots (x_n + y_n)$ is computable by a width-2 ROABP in the order $(x_1, y_1, x_2, y_2, \ldots, x_n, y_n)$; but any ROABP that reads all the $\mathbf{x}$-variables before the $\mathbf{y}$-variables (e.g. in the order $(x_1, \ldots, x_n, y_1, \ldots, y_n)$) requires width $2^{\Omega(n)}$. The existence of such polynomials naturally leads to the following classes of polynomials (defined in Section 2).

- ROABP$[\exists](n, d, w)$ - n-variate, individual degree d polynomials that are computable by a width-w ROABP in *some* order $\sigma \in S_n$.
- ROABP$[\forall](n, d, w)$ - n-variate, individual degree d polynomials that are computable by a width-w ROABP in *every* order.

Clearly, ROABP$[\forall](n, d, w) \subseteq$ ROABP$[\exists](n, d, w)$, and the former class requires exponential width to simulate the latter, due to the example discussed above.

Observe that an ROABP in the order id $= (x_1, \ldots, x_n)$, can be written as $\mathbf{u}^\intercal \cdot M_1(x_1) \cdot M_2(x_2) \cdots M_n(x_n) \cdot \mathbf{v}$, with entries of each M_i being univariate polynomials in $\mathbb{C}[x_i]$. Alternatively, we can view the same, as $\mathbf{u}^\intercal \left(\prod_{i \in [n]} \left(A_{i,0} + A_{i,1} x_i + \cdots + A_{i,d} x_i^d \right) \right) \mathbf{v}$, by interpreting each M_i as a univariate with matrices as coefficients. We refer to these matrices $\{A_{i,j}\}$ as the *coefficient matrices* of the ROABP.

Now based on the properties of the coefficient matrices $\{A_{i,j}\}$, one can define the following models and the corresponding classes.

- *Commutative ROABPs:* ROABPs where all the $n(d+1)$ coefficient matrices commute with each other (see Definition 12).
  $\mathsf{commROABP}(n, d, w)$ - n-variate, individual degree d polynomials that are computable by a width w commutative ROABP.
- *Diagonal ROABPs:* ROABPs where all the $n(d+1)$ coefficient matrices are diagonal matrices (see Definition 13).
  $\mathsf{diagROABP}(n, d, w)$ - n-variate, individual degree d polynomials that are computable by a width w diagonal ROABP.

First of all, $\mathsf{commROABP}(n, d, w) \subseteq$ ROABP$[\forall](n, d, w)$ for any n, d, w, since the coefficient matrices in any commutative ROABP are commutative, and one can multiply the matrices in any order to get the same result. Likewise, as all diagonal matrices commute with each other, $\mathsf{diagROABP}(n, d, w) \subseteq \mathsf{commROABP}(n, d, w)$. In this paper, we investigate commutative and diagonal ROABPs to understand if and when these two classes are the essentially (up to polynomial-factors) equal.

While it is indeed true that even diagonal ROABPs are universal, it is reasonable to ask if there are any interesting polynomial families that are efficiently computable by commutative and diagonal ROABPs. In this regard, let us begin by looking at the constructions of "all-order-ROABPs" for two well studied polynomial families: *elementary symmetric polynomials* and *powers of linear forms*. Incidentally, these constructions can naturally be interpreted as commutative ROABPs, and further, they even lead to diagonal ROABPs that achieve the best known upper bounds. We believe that these examples should serve as an additional motivation to study the models of commutative and diagonal ROABPs.

▶ **Definition 1** (Elementary Symmetric Polynomials). *The n-variate elementary symmetric polynomial of degree d, denoted by ESym_n^d is defined as follows.*

$$\mathsf{ESym}_n^d(\mathbf{x}) := \sum_{\substack{S \subset [n] \\ |S| = d}} \prod_{i \in S} x_i \tag{1.1}$$

Following is a folklore construction (with a minor tweak) of an ROABP for ESym_n^d which is provably tight owing to the characterisation result by Nisan [14] (see full version). We illustrate the construction for $n = 5$ and $d = 3$ in the full version, and give the general recipe here without a proof of correctness.

▶ **Construction 1.2.** *For any $n, d \in \mathbb{N}$ such that $d \leq n$, we have the following.*

$$\mathsf{ESym}_n^d(\mathbf{x}) = (M(x_1)M(x_2)\cdots M(x_n))\,[1, d+1],$$

where for all i, $M(x_i)$ is a $(d+1) \times (d+1)$ matrix such that $M(x_i)[k,k] = 1$ for all $1 \leq k \leq (d+1)$, and $M(x_i)[k,k+1] = 1$ for all $1 \leq k \leq d$; all other entries of $M(x_i)$ are zero.

The matrix $M(x_i)$ can also be written as $(I + Ax_i)$, where A is a matrix with 1s on its super-diagonal and zeros everywhere else, and I is the identity matrix. This gives the expression: $\mathsf{ESym}_n^d(\mathbf{x}) = ((I + Ax_1)(I + Ax_2)\cdots(I + Ax_n))_{(1,d+1)} = \mathbf{u}^\intercal \left(\prod_{i\in[n]}(I + Ax_i)\right)\mathbf{v}$, for the obvious choice of $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{(d+1)}$.

We can now make the following sequence of simple observations about this construction.

- All the coefficient matrices of the above ROABP: I and A, commute with each other. Thus, it is a commutative ROABP.
- $(I+Ax_1)(I+Ax_2)\cdots(I+Ax_n) = \sum_{0\leq j\leq n} \mathsf{ESym}_n^j\, A^j = \sum_{0\leq j\leq d} \mathsf{ESym}_n^j\, A^j$, since $A^j = 0$ for all $j \geq (d+1)$.
- For every $0 \leq j \leq d$, only the jth power of A that has a 1 in the $(1, 1+j)$th entry. Therefore, the $(1, d+1)$th entry of $(I + Ax_1)(I + Ax_2)\cdots(I + Ax_n)$ exactly computes the coefficient of A^d, which is ESym_n^d.

This perspective along with elementary interpolation, then leads us to the following *depth-3-multilinear* circuit for ESym_n^d of *top fan-in* $(n+1)$ for all values of d, that is attributed to Ben-Or ([21]). This also happens to give the following nearly-optimal construction for a diagonal ROABP computing ESym_n^d.

▶ **Construction 1.3.** *For any $n, d \in \mathbb{N}$ and distinct $a_0, a_1, \ldots, a_n \in \mathbb{F}$, there exist constants $\beta_0, \beta_1, \ldots, \beta_n \in \mathbb{F}$ such that*

$$\mathsf{ESym}_n^d(\mathbf{x}) = \sum_{0\leq j\leq n} \beta_j(1 + a_j x_1)(1 + a_j x_2)\cdots(1 + a_j x_n)$$

Just as the commutative ROABP for $\mathsf{ESym}_n^d(\mathbf{x})$ leads us to Ben-or's construction of a diagonal ROABP, we also observe that the commutative ROABP computing dth power of an n-variate linear form $(x_1 + x_2 + \cdots + x_n)^d$ gives us the *duality trick* of Saxena [18] (see e.g. [19, Lemma 17.13]). We refer the interested reader to the full version.

As the coefficient matrices of diagonal ROABPs are diagonal matrices it is not difficult to observe that they are exactly *sums-of-products-of-univariates*. Thus, from the duality trick, we observe that diagonal ROABPs can efficiently simulate *diagonal depth* 3 *circuits* (a.k.a. *depth*-3 *powering circuits*) denoted by $\Sigma\wedge\Sigma$. That is, $\Sigma\wedge\Sigma(n, d, s) \subseteq \mathsf{diagROABP}(n, d, O(n, d, s))$.Also, a separation between these two classes is known due to the exponential lower bound from [15] for $x_1 \ldots x_n$ against the model $\Sigma \wedge \Sigma$. In essence, we have the following containments between classes[2], where each $\mathcal{C}$ stands for the class of n-variate, degree-d polynomials whose $\mathcal{C}$-size is $\mathrm{poly}(n, d)$.

$$\Sigma \wedge \Sigma \subsetneq \mathsf{diagROABP} \subseteq \mathsf{commROABP} \subseteq \mathrm{ROABP}[\forall] \subsetneq \mathrm{ROABP}[\exists]$$

[2] We have more intricate relationships between classes concerning ROABPs. See Subsection 1.3

Looking at the above hierarchy, we firstly realise that nearly optimal separations are known at the two "extremes", but nothing is known about the intermediate levels. Further, since the intermediate levels are far more algebraically structured (coefficient matrices arising from special commutative algebras), it is reasonable to expect finer separations for these classes. Unfortunately, all the lower bounds that we know for diagonal and commutative ROABPs are those that are known for ROABP[$\forall$].

Secondly, even though diagonal ROABPs (*sum-of-products-of-univariates*) may be of independent interest as they subsume $\Sigma \wedge \Sigma$ circuits, they are also interesting from the point of view of polynomial identity testing. Owing to the algebraic structure of their coefficients, one can expect efficient PIT algorithms for these classes. But again, the best PIT algorithms that we know for diagonal and commutative ROABPs are those we know for ROABP[$\forall$]. We discuss more about polynomial identity testing algorithms for these classes in the full version.

1.1 Our Results

We now move to the central questions addressed in this article. In particular, we wish to understand if the classes commROABP and diagROABP are equal up to polynomial factors; this can be more formally stated as follows.

▶ **Question 1.4.** *Given an n-variate, individual degree d polynomial $f(\mathbf{x})$ computable by a width w commutative ROABP(i.e. $f \in \mathsf{commROABP}(n,d,w)$), does there exist a diagonal ROABP computing f of width* $\mathrm{poly}(n,d,w)$*?*

A measure that is often used to prove lower bounds against structured models (e.g. almost every lower bound against $\Sigma \wedge \Sigma$, and more recently [11]) is the *dimension of partial derivatives*, a complexity measure which was introduced by Nisan and Wigderson [15] (see Definition 16). For any polynomial $f \in \mathbb{C}[\mathbf{x}]$, the partial derivative complexity of f (denoted by $\mathrm{DPD}(f)$) is the dimension of the space spanned by *all* the partial derivatives of f. Nisan and Wigderson [15] observed that any n-variate, degree d polynomial $f(\mathbf{x})$ that has a $\Sigma \wedge \Sigma$ circuit of size s has $\mathrm{DPD}(f) \leq s(d+1)$. Therefore it is natural to ask whether the $\Sigma \wedge \Sigma$-size of every polynomial f is polynomially related to its dimension of partial derivatives. We formalize this question as follows.

▶ **Question 1.5.** *Does there exist a* constant *c such that for any n-variate, degree-d polynomial $f(\mathbf{x})$ with $\mathrm{DPD}(f) \leq s$, we have that the smallest $\Sigma \wedge \Sigma$ circuit that computes $f(\mathbf{x})$ has size at most $(nds)^c$?*

The size of the smallest $\Sigma \wedge \Sigma$ circuit for a polynomial is a well studied notion called the *Waring rank of f* (denoted by $\mathrm{WR}(f)$). Question 1.5 essentially asks if the Waring rank and the dimension partial derivatives of a polynomial are same up to polynomial factors. Unfortunately, at the moment we do not know the answers to either Question 1.4 or Question 1.5. However, our main result gives a rather surprising connection between Question 1.4 and Question 1.5. Specifically, we show that an positive answer to Question 1.5 answers Question 1.4 in the affirmative!

▶ **Theorem 2.** *For any $n, r \in \mathbb{N}$, let $S(r,m)$ denote the smallest $\Sigma \wedge \Sigma$-size required to compute any r-variate polynomial f with $\mathrm{DPD}(f) \leq m$.*
Then for all $n, d, w \in \mathbb{N}$, $\mathsf{commROABP}(n,d,w) \subseteq \mathsf{diagROABP}\left(n,d,S(w^2,w^2)nw^4\right)$.

▶ Remark 3. In fact, it can be inferred from our proof that a super-polynomial separation between commROABP and diagROABP will yield an explicit polynomial that witnesses a super-polynomial separation between dimension of partial derivatives and Waring rank. We elaborate on this in Remark 22.

A different (and perhaps equally surprising) consequence of Theorem 2 is that a super-polynomial separation between commutative ROABPs and diagonal ROABP will also give a super-polynomial separation dimension of partial derivatives and Waring rank. Note that not only do we not know the answers to Question 1.4 or Question 1.5, it is somewhat frustrating that we do not even know of a candidate polynomial that could potentially separate these classes. We expect that our analysis of these models that goes into proving the result above could help in making some progress in either of these questions.

1.2 An overview of the proof

We start by asking when diagonal ROABPs can efficiently simulate commutative ROABPs. This question naturally leads us to study properties of matrices that commute with each other. In particular, we analyse *commutative rings* generated by matrices that commute with each other.

A very high level overview. The results of Marinari, Möller, Mora [12], and Möller and Stetter [13] provide a characterisation of commutative rings of $w \times w$ matrices in terms of polynomials whose *dimension of partial derivatives* is at most $\text{poly}(w)$. In the special case when these matrices are all diagonal, the same polynomials happen to have *Waring rank* at most w. Further, we observe that if the polynomials corresponding to a n-variate, width-w commutative ROABP have *Waring rank* at most s, then it can be simulated by a diagonal ROABP of width $\text{poly}(n, w, s)$. This is essentially our main result. We now explain the characterisation given by [12] and [13] in a bit more detail.

Characterising rings of matrices

Consider the ring generated by a $w \times w$ matrix A, given by $\mathbb{C}[A] := \{q(A) : q(t) \in \mathbb{C}[t]\}$. The ring has at most w *linearly independent* matrices, as the characteristic polynomial of A gives a way to express A^w as a linear combination of lower powers of A. In fact, the ring $\mathbb{C}[A]$ is characterised by the *ideal* of all polynomials that are divisible by the *minimal polynomial of* A (see Fact A.1). This characterisation has an appropriate analogue for general matrix rings, as follows.

Suppose that $A_1, \ldots, A_r \in \mathbb{C}^{w \times w}$ commute with each other, and let $\mathbb{C}[A_1, \ldots, A_r]$, defined as $\{g(A_1, \ldots, A_r) : g(\mathbf{t}) \in \mathbb{C}[\mathbf{t}]\}$, be the ring generated by them[3]. Analogous to the univariate (singly-generated) case, we then consider the *ideal of dependencies* for the matrices $A_1, \ldots, A_r$: $J = \{p(\mathbf{t}) \in \mathbb{C}[\mathbf{t}] : p(A_1, \ldots, A_r) = 0\}$. As it turns out, $\mathbb{C}[A_1, \ldots, A_r]$ is indeed characterised by the ideal J (see Lemma 23).

Before delving further into the ideal of dependencies, we remark a structural property of polynomials that admit a diagonal ROABP of a certain width.

Understanding diagonal ROABPs. Consider the diagonal ROABP (depth-3 multilinear circuit) for the *elementary symmetric polynomial* $\mathsf{ESym}_{n,d}$ that is attributed to Ben-Or (see e.g. [21]). One first constructs the polynomial $g(t, \mathbf{x}) := (1 + tx_1)(1 + tx_2) \cdots (1 + tx_n)$, and then obtains $\mathsf{ESym}_{n,d}$ as the coefficient of t^d in $g(t, \mathbf{x})$, using interpolation. It turns out that any diagonal ROABP computing a polynomial $f(\mathbf{x})$ can similarly be seen as expressing f as a linear combination of evaluations of a *low-degree* $g(t, \mathbf{x})$ that is a "product of univariates"

[3] Any ring of $w \times w$ matrices is generated by at most w^2 matrices.

(see Observation 18). Here, the number of evaluations needed is *equal* to the width of the ROABP. Moreover the converse of this statement is also true, thus giving us an equivalent formulation for diagonal ROABPs.

Therefore, we analyse the ideal J with the goal of expressing the corresponding commutative ROABP as a *sum of* $\mathbf{t}$*-evaluations* of some $G(\mathbf{t}, \mathbf{x}) = G_1(\mathbf{t}, x_1) \cdot G_2(\mathbf{t}, x_2) \cdots G_n(\mathbf{t}, x_n)$.

The ideal of dependencies. Let us first make our statement about $\mathbb{C}[A_1, \ldots, A_r]$ being characterised by J a bit more precise: there is a *ring-isomorphism* between $\mathbb{C}[A_1, \ldots, A_r]$ and the *quotient ring* $\mathbb{C}[\mathbf{t}]/J$. Therefore it is crucial to understand J (and $\mathbb{C}[\mathbf{t}]/J$) to understand the ring of matrices, in order to move towards the above mentioned goal.

Let $p(t)$ be the minimal polynomial of some matrix A, and consider the ideal $\langle p \rangle$. If $p(t) = (t-5)^3$, then we know that any $q(t)$ belongs to $\langle p \rangle$ *if and only if* the first 3 derivatives of $q(t)$ vanish at $t = 5$; i.e. $q(5) = q'(5) = q''(5) = 0$. In general, for $p(t) = (t-a_1)^{e_1}(t-a_2)^{e_2} \cdots (t-a_k)^{e_k}$, membership in the ideal $\langle p \rangle$ is *characterised* by the first e_i derivatives vanishing at $t = a_i$, for *each* $i = 1, 2, \ldots, k$. Moreover, the polynomial "$q(t)$ mod $p(t)$" can be obtained by applying a *linear transformation* on the evaluations of the $e_1, \ldots, e_k$ derivatives at the respective points $a_1, \ldots, a_k$.

We now extend this understanding to the multivariate setting. We already have the correct analogue for $\langle p \rangle$, which we call the ideal of dependencies J. Next, we need a characterisation for "$g(\mathbf{t})$ mod J" in terms of some derivatives of $g(\mathbf{t})$ evaluated at some points related to J. While these choices were quite clear in the univariate setting from p; the multivariate setting requires a little more care. Fortunately for us, the works of Marinari, Möller, Mora [12], and Möller and Stetter [13] provide an adequate solution.

Firstly, observe that J has a finite *variety* (common zeroes of all polynomials in J). Thus the variety $\mathbf{V}(J)$ is a good multivariate analogue for the set of evaluation points. The other ingredient that we require is a compatible notion of "multiplicity of J" at a point $\bar{\alpha}$ in its variety. For this, [12] look at the set of all *partial derivative operators* (see Definition 25) which map *every* polynomial in J to a polynomial that vanishes at $\bar{\alpha}$. These operators form a vector space over $\mathbb{C}$, and the "multiplicity of J at $\bar{\alpha}$" is then defined as the *dimension* of this vector space.

In the univariate setting, the multiplicity of q at a point a_i is defined as the *highest* number e_i such that the first e_i derivatives of q vanish at the point a_i. Thus, one can naturally identify a "highest derivative", with the other derivatives being its "down-shifted versions". Analogously, the derivative operator space corresponding to J and a point $\mathbf{v} \in \mathbf{V}(J)$ is *closed under taking down-shifts* (see Definition 29). An ideal J with $\mathbf{V}(J) = \{\bar{\alpha}_1, \ldots, \bar{\alpha}_k\}$, is then captured by a collection of z vector spaces of derivative operators $\Delta_1, \Delta_2, \ldots, \Delta_k$, in the following sense (see Lemma 32).

- For each $i \in [k]$, Δ_i corresponds to the point $\bar{\alpha}_i$ and is down-closed.
- Dimension of the quotient ring $\mathbb{C}[\mathbf{t}]/J$ is $w = \dim(\Delta_1) + \dim(\Delta_2) + \cdots + \dim(\Delta_k)$.
- Let $\{D_{i,1}, \ldots, D_{i,w_i}\}$ be a basis of Δ_i. Then there exists a map $\Phi : \mathbb{C}^w \to \mathbb{C}[\mathbf{t}]/J$ such that for any polynomial $q(\mathbf{t})$, Φ maps the w values: $\{D_{i,j}(q)(\mathbf{v}_i)\}$, to the "remainder polynomial" ($q(\mathbf{t})$ mod I).

Further, Möller and Stetter [13] show that the map Φ stated above is just a linear transformation (see Lemma 36).

Consequences for ROABPs. We now outline the proof of our main result (Theorem 2).

- Given a commutative ROABP $f(\mathbf{x}) = \mathbf{b}^\intercal \cdot \prod_{i \in [n]} \left(A_{i,0} + A_{i,1}x_i + \cdots + A_{i,d}x_i^d\right) \cdot \mathbf{c}$ of width w, we define $F(\mathbf{x}) := \prod_{i \in [n]} \left(A_{i,0} + A_{i,1}x_i + \cdots + A_{i,d}x_i^d\right)$ to be a matrix of polynomials. Then, $f(\mathbf{x})$ is just a linear combination (given by $\mathbf{b}\mathbf{c}^\intercal$) of the entries of F.

- We then identify a set of matrices $A_1, \ldots, A_r$ that generate the coefficient-matrix-ring; i.e. $\mathbb{C}[A_1, \ldots, A_r] = \mathbb{C}[A_{1,0}, \ldots, A_{1,d}, \ldots, A_{n,d}]$. As we can always use the coefficient matrices themselves, and because we are dealing with $w \times w$ matrices, $r \leq \min(w^2, n(d+1))$.
- Let J be the ideal of dependencies for $A_1, \ldots, A_r$ and suppose the normal set of J (see Definition 35) has size, say $m \leq w^2$. Then each $A_{i,j}$ is a polynomial in $A_1, \ldots, A_r$ that has $\leq m$ monomials.
- For each i, j, suppose $G_{i,j}(t_1, \ldots, t_r)$ is the polynomial such that $G_{i,j}(A_1, \ldots, A_r) = A_{i,j}$; the entries of $A_{i,j}$ are linear combinations of $\mathbf{t}$-coefficients of $G_{i,j} = (G_{i,j} \bmod J)$. Then we observe that $G(\mathbf{t}, \mathbf{x}) := \prod_{i \in [n]} \left(G_{i,0}(\mathbf{t}) + G_{i,1}(\mathbf{t})x_i + \cdots + G_{i,d}(\mathbf{t})x_i^d\right)$, such that $G(A_1, \ldots, A_r, \mathbf{x}) = F(\mathbf{x})$. This means that even $f(\mathbf{x})$ is a linear combination of the $\mathbf{t}$-coefficients of $(G(\mathbf{t}, \mathbf{x}) \bmod J)$, as it is a linear combination of the entries of $F(\mathbf{x})$. We prove this in Lemma 20.
- Now let $\mathbf{V}(J) = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and for each $\ell \in [k]$ let $\{D_{\ell,1}, \ldots, D_{\ell,m_\ell}\}$ be a basis for the derivative operator space corresponding to $\mathbf{v}_\ell$. Then from the results of [12, 13] we get that for any $g(\mathbf{t})$, every $\mathbf{t}$-coefficient of $(g(\mathbf{t}) \bmod J)$ is a fixed linear combination of the m values given by $(D_{\ell,*}(g))(v_\ell)$.
- This brings us one step away from our goal of expressing $f(\mathbf{x})$ as a linear combination of $\mathbf{t}$-evaluations of some $G(\mathbf{t}, \mathbf{x})$ which is a product of univariates. What we need is a way to express each of $(D_{\ell,*}(G))(\mathbf{v})$ as a linear combination of $\mathbf{t}$-evaluations of $G(\mathbf{t}, \mathbf{x})$.
- It turns out that the number of evaluations of $G(\mathbf{t}, \mathbf{x})$ required to compute $(D_{\ell,*}(G))(\mathbf{v})$ is $\mathrm{poly}(\deg(h_{\ell,*}), \mathrm{WR}(h_{\ell,*}))$, where $h_{\ell,*}$ is the *polynomial corresponding to* $D_{\ell,*}$ (see paragraph below Definition 25). This is a non-trivial fact; we prove it in Lemma 21.
- Finally, since each space Δ_ℓ is *down-closed*, we have that the dimension of partial derivatives $\mathrm{DPD}(h_{\ell,*}) \leq \dim(\Delta_\ell) \leq m$ for each $h_{\ell,*}$. Therefore, using the hypothesis that $\mathrm{WR}(h) = \mathrm{poly}(r, \mathrm{DPD}(h))$ for any r-variate h, we get that $(D_{\ell,*}(G))(\mathbf{v})$ can be expressed as a linear combination of $\mathrm{poly}(r, \mathrm{DPD}(h_{\ell,*}), \deg(h_{\ell,*})) = \mathrm{poly}(r, m)$ evaluations of $G(\mathbf{t}, \mathbf{x})$ for each $D_{\ell,*}$.
- Combining all the above observations, we can see that the hypothesis implies that $f(\mathbf{x})$ can indeed be written as a linear combination of $\mathrm{poly}(r, m) = \mathrm{poly}(n, d, w)$ evaluations of $G(\mathbf{t}, \mathbf{x})$, thereby proving Theorem 2.

1.3 Landscape of ROABP classes

As mentioned earlier, although Theorem 2 relates Question 1.4 and Question 1.5, the answer to both these questions remain unknown. In this regard, we would like to conjecture that the answer to both questions is false.

▶ **Conjecture 4.** *There exists an explicit n-variate degree d polynomial $f(\mathbf{x})$ such that $f \in \mathsf{commROABP}(n, d, \mathrm{poly}(n, d))$ and any diagonal ROABP computing f requires width $n^{\omega(1)}$.*

▶ **Conjecture 5.** *There exists an explicit n-variate polynomial $f(\mathbf{x})$ of degree $\mathrm{poly}(n)$ such that $\mathrm{DPD}(f) = \mathrm{poly}(n)$ but $\mathrm{WR}(f) = n^{\omega(1)}$.*

Even though many would agree that Conjecture 4 and Conjecture 5 are probably true, we do not even know of any candidate polynomial that will witness the truth of this conjecture. In relation to this, we remark that the following statement can be inferred from our proof of Theorem 2. If there exists a commutative ROABP of width $\mathrm{poly}(n, d)$ computing an n-variate, degree-d polynomial f, which requires diagonal ROABPs of super-polynomial width, then the commutative ROABP for f gives a different explicit polynomial h that has

polynomial dimension of partial derivatives, but has super-polynomial Waring rank (see Remark 22 for details). As a result, even a candidate polynomial for proving Conjecture 4 remains unknown.

In the context of Conjecture 4, we remark the following connection between diagonal ROABPs and tensor rank.

▶ Remark 6. Observe that the width of a diagonal ROABP exactly captures the *tensor rank* of the corresponding *tensor*. A tensor $T : [d]^n \to \mathbb{C}$ of order[4] n can naturally be viewed as a polynomial $f_T = \sum_{\mathbf{i}\in[d]^n} T(i_1, \ldots, i_n) x_1^{i_1} \cdots x_n^{i_n}$. The (tensor) rank of any T (denoted by $\mathsf{TR}(T)$) is the smallest r such that T can be expressed as sum of r elementary tensors. Thus for any tensor T, $\mathsf{TR}(T) = r$ if and only if $f_T(\mathbf{x})$ can be expressed as sum of r many products of univariates; which immediately implies $\mathsf{diagROABP}(n, d, w) = \{f_T \in \mathbb{C}[\mathbf{x}] \mid \mathsf{TR}(T) \leq w\}$. Obtaining strong lower bounds on the rank of explicit tensors is a major open problem in algebraic complexity theory, where the goal is to obtain an explicit tensor T of order-n such that $\mathsf{TR}(T) = d^{n(1-o(1))}$ (see e.g. [17]).

Remark 6 tells us that proving strong width lower bounds against diagonal ROABPs could potentially imply lower bounds on the rank of explicit tensors. While this could partially explain why there are no separations between diagonal ROABPs and commutative or "all-order" ROABPs, it is also worth mentioning that order-n tensors for a *growing parameter* n are rarely studied in the context of tensor rank lower bounds.

With regard to Conjecture 5, we briefly discuss some known results about the problem of computing the dimension of the partial derivative space.

Shitov [20] showed that given any degree 3 polynomial f in its sparse representation, computing $\mathrm{WR}(f)$ is NP-hard, by reducing it to computing the tensor rank of order 3 *symmetric tensors*. On the other hand, when a polynomial f is presented in its sparse representation (as sum of monomials), García-Marco, Koiran, Pecatte and Thomassé [7] prove that computing the dimension of the partial derivative space is #P-hard (not known to be #P-complete). Thus, even though computing Waring rank is a hard problem, it is not quite clear if disproving Conjecture 5 goes against it. Moreover, it is possible that Waring rank is easy to *approximate* up to polynomial factors, which is all that a disproof of Conjecture 5 would imply. On a related note, Kayal [9] gave a randomised $\mathrm{poly}(n, d)$-time algorithm to compute the waring rank of an n-variate, degree-d polynomial that is given as a blackbox (in the *non-degenerate case*).

Although the results in this article entirely concern Question 1.5 and Question 1.4, there are several other interesting open questions surrounding the landscape of complexity classes involving ROABPs. We discuss these interconnections between ROABP classes now, and later illustrate them in Figure 1.

Let us consider the class of polynomials computed by ROABPs that remain unchanged by interchanging layers in the branching program[5]. We prefer to use the term *layer-commutative ROABPs* (denoted by $\mathsf{layer\text{-}commROABP}(n, d, w)$) to denote the class of n-variate degree d polynomials computed by an ROABPs such that if $f = u^T M_1(x_1) \cdots M_n(x_n) v$ then the matrices of univariate polynomials $M_1, \ldots, M_n$ commute. That is, $M_i(x_i)M_j(x_j) = M_j(x_j)M_i(x_i)$ for all $i, j \in [n]$. Clearly, $\mathsf{layer\text{-}commROABP}(n, d, w) \subseteq \mathrm{ROABP}[\forall](n, d, w)$, and $\mathsf{commROABP}(n, d, w) \subseteq \mathsf{layer\text{-}commROABP}(n, d, w)$. This immediately leads us to the following two open questions whose answer seems unclear at the moment.

[4] Commonly used term in the literature about tensors; not be confused with the order of an ROABP.

[5] The class $\mathrm{ROABP}[\forall](n, d, w)$ has been studied in the context of PIT, and is sometimes called *commutative ROABPs* in some works (e.g. [8]). We use a different notation to avoid any ambiguity.

▶ **Question 1.6.**

1. *Are* layer-commROABP(n, d, w) *and* ROABP$[\forall](n, d, w)$ *equivalent up to a polynomial blow-up in the width* w*?*
2. *Are* commROABP(n, d, w) *and* layer-commROABP(n, d, w) *equivalent up to a polynomial blow-up in the width* w*?*

We hope that a better understanding the algebra associated with commutative ROABPs may shed light on the answers to above questions.

Along with the complexity of computing polynomials exactly, another notion that is considered in algebraic complexity theory and more specifically in *geometric complexity theory*, is *border complexity* of polynomials. Let $\mathcal{C}$ be a class of polynomials. We say that f is in the class $\overline{C}$ (border of $\mathcal{C}$), if f can be "arbitrarily-approximated" by a circuit in $\mathcal{C}$. That is, there exists a polynomial $g(\epsilon) \in \mathbb{C}(\epsilon)$ in class $\mathcal{C}$ such that $f = \lim_{\epsilon \to 0} g$. The *border-complexity* of f is then at most the size of the circuit computing g. Clearly, $\mathcal{C} \subseteq \overline{\mathcal{C}}$. Understanding whether $\mathcal{C} = \overline{\mathcal{C}}$ for interesting classes such as VP and VBP are major open problems in algebraic complexity theory. Here, we are interested in the case when $\mathcal{C} =$ diagROABP(n, d, w) (defined in Definition 17).

▶ **Question 1.7.** *Is there a super-polynomial separation between the classes* diagROABP(n, d, w) *and* $\overline{\text{diagROABP}}(n, d, w)$*?*

As diagROABP$(n, d, w) = \{f_T \in \mathbb{C}[\mathbf{x}] \mid \text{TR}(T) \leq w\}$, we have $\overline{\text{diagROABP}(n, d, w)} = \{f_T \in \mathbb{C}[\mathbf{x}] \mid \underline{\text{TR}}_{\mathbb{C}}(f) \leq w\}$. Here, $\underline{\text{TR}}(f)$ denotes the border rank of tensors. Border rank of tensors is studied extensively in several contexts for instance, border rank of *matrix multiplication tensor* is used to obtain bounds on the arithmetic complexity of matrix multiplication. In this setting, the order of the tensor is usually bounded by a constant, and this setting slightly deviates from the main theme algebraic circuit complexity.

It can be checked that just like commROABP(n, d, w), $\overline{\text{diagROABP}}(n, d, w)$ is also contained in ROABP$[\forall](n, d, w)$ (because ROABP-complexity is characterised by rank, which is a continuous measure). However, it is unclear if these two ways of "generalising" diagonal ROABPs have different computational powers. This brings us to the following question.

▶ **Question 1.8.** *Are the classes* $\overline{\text{diagROABP}}(n, d, w)$ *and* commROABP(n, d, w) *equivalent up to polynomial factors?*

Note that Question 1.8 is linked to the question of understanding commROABP(n, d, w) and ROABP$[\forall](n, d, w)$ in Question 1.6. Also, answering this question in the affirmative is similar in spirit to the recent "de-bordering" results due to Dutta et al. [5]. They proved that the border of constant *top fan-in* depth three circuits is contained in the class VBP. Here, Question 1.8 is essentially asking if for the class of diagonal ROABPs (albeit with unbounded fan-in), the border is contained in a much simpler class of commutative ROABPs? However, answering this in the negative could potentially be as hard as (or even harder than) separating commutative ROABPs from diagonal ROABPs. In fact, it is not even clear if these two classes should be comparable (contained in one another). We believe that any answer to Question 1.8 would be an interesting development in algebraic complexity theory.

We summarize all the models and the interconnections between the structured ROABP classes in Figure 1.

2 Preliminaries

We now formally define the classes of polynomials and other algebraic models of computation that we study in this paper. First, we fix some notation.

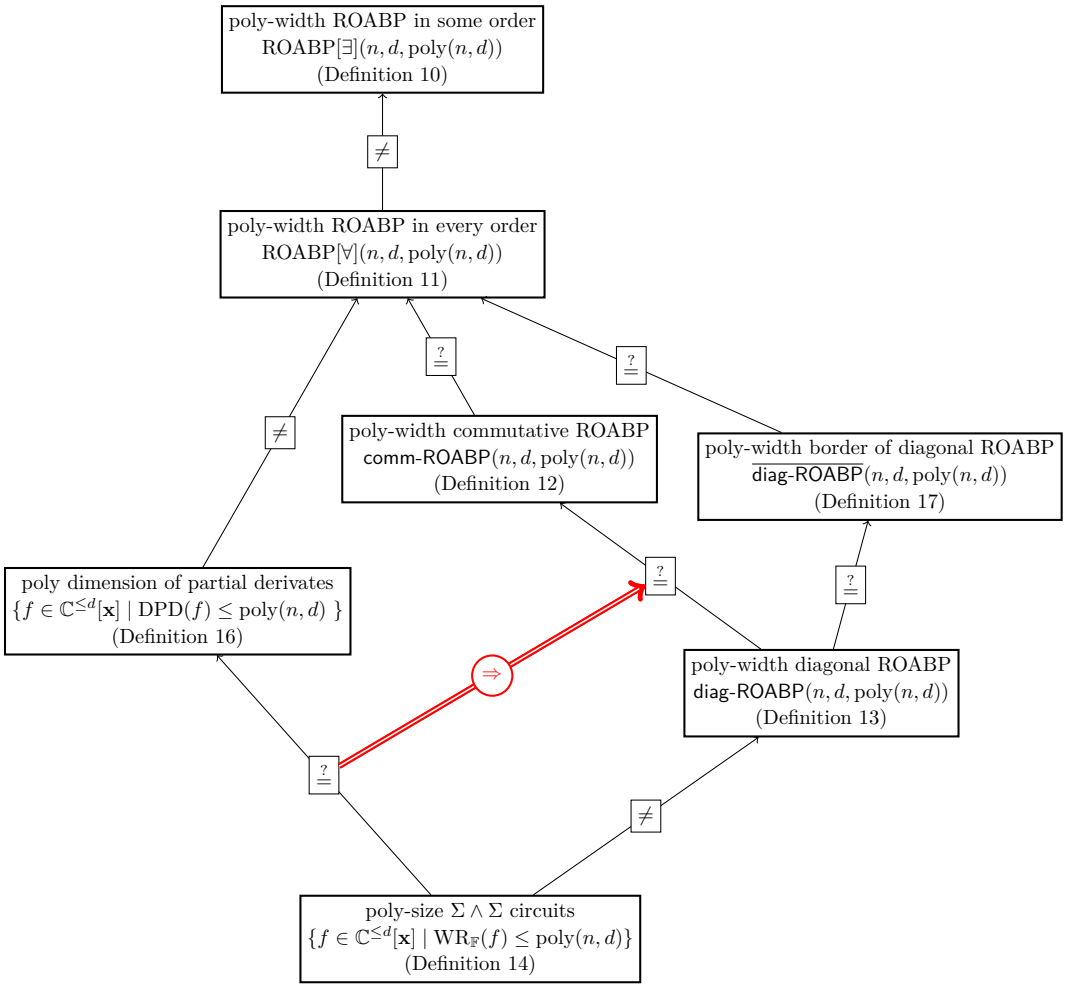

Figure 1 The ROABP landscape: edges denote bottom-up inclusion, Theorem 2 is in red.

Notation

- We use the shorthand $[n]$ to denote the set $\{1, 2, \ldots, n\}$.
- We use boldface letters like $\mathbf{x}, \mathbf{t}, \mathbf{A}, \mathbf{b}$, to denote sets/vectors. The individual elements/-coordinates are denoted by indexed versions of the same characters: $\mathbf{A} = \{A_1, \ldots, A_r\}$. Whenever the size of these sets is not clear from context, we denote them using subscripts: $\mathbf{x}_{[n]} = \{x_1, \ldots, x_n\}$.
- For a polynomial $f(\mathbf{x})$, we denote *support of* f the set of monomials appearing in f with a nonzero coefficient by $\text{supp}(f)$.
- For $\mathbf{x} = \{x_1, \ldots, x_n\}$, and any vector $\mathbf{e} \in \mathbb{N}^n$, we use the shorthand $\mathbf{x}^{\mathbf{e}}$ to denote the monomial $x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}$.
- For a polynomial $f(\mathbf{x})$ and a monomial $\mathbf{x}^{\mathbf{e}}$, we use $\partial_{\mathbf{e}} f$ to denote the partial derivative $\frac{\partial^{|\mathbf{e}|} f}{\partial x_1^{e_1} \ldots \partial x_n^{e_n}}$.
- For a matrix M, $M[i, j]$ denotes its (i, j)th entry.

We start by defining *algebraic circuits* and *algebraic branching programs*. Note that we work with the field of complex numbers unless mentioned otherwise.

▶ **Definition 7** (Algebraic circuits). *An* algebraic circuit *is specified by a directed acyclic graph, with leaves (in-degree zero; also called* inputs*) labelled by field constants or variables, and internal nodes labelled by $+$ or $\times$. The nodes with out-degree zero are called the* outputs *of the circuit. Computation proceeds in the natural way, where inductively each $+$ gate computes the sum of its children and each $\times$ gate computes the product of its children.*

The size *of the circuit is defined as the number of nodes in the underlying graph.*

▶ **Definition 8** (Algebraic Branching Programs). *An algebraic branching program is a layered, directed graph. There are two special vertices, source s and sink t which are the only vertices in the first and last layers, respectively. All the edges in the graph are from one layer to the consecutive layer. Each edge is labelled by a univariate polynomial in the underlying variables over the underlying field. Each path from s to t computes the product of the edge labels and the ABP computes the sum of all the paths from s to t. Then, any ABP can be viewed as a product of matrices (each matrix having univariate polynomials as its entries) and the ABP computes the $(1,1)$th entry of the matrix product. The maximum number of vertices in a single layer (dimension of the largest matrix in the product) is called its* width*. The size of the ABP is the total number of vertices in it.*

We now define the various structured ROABPs and other related classes that are the main objects of interest in our paper. We start by defining the basic model of ROABPs.

▶ **Definition 9** (Read-once Oblivious ABPs). *Over the field $\mathbb{C}$ of complex numbers, a* read-once oblivious algebraic branching program *or an* ROABP*, computes an n-variate,* individual *degree d polynomial using a matrix-vector product of the following form.*

$$R(\mathbf{x}) = \mathbf{u}^{\intercal} \cdot M_1(x_{\sigma(1)}) \cdot M_2(x_{\sigma(2)}) \cdots M_n(x_{\sigma(n)}) \cdot \mathbf{v}$$

where

- *For each $i \in [n]$, the matrix $M_i(x_{\sigma(i)})$ has entries that are univariates of degree $\leq d$ in the variable $x_{\sigma(i)}$,*
- *$\mathbf{u} \in \mathbb{C}^{w_0}$, $M_1(x_{\sigma(1)}) \in (\mathbb{C}[x_{\sigma(1)}])^{w_0 \times w_1}$, …, $M_i(x_{\sigma(i)}) \in (\mathbb{C}[x_{\sigma(i)}])^{w_i \times w_{i+1}}$, …, $\mathbf{v} \in \mathbb{C}^{w_n}$,*
- *the* width *w of the ROABP R is defined as $w = \max\{w_0, w_1, \ldots, w_n\}$,*
- *the permutation σ is called as* the order of the ROABP *R.*

The following two subclasses of polynomials then follow naturally from the definition of ROABPs.

▶ **Definition 10** (ROABPs in some order). *For $n, d, w \in \mathbb{N}$, an n-variate polynomial $f(\mathbf{x})$ of* individual *degree d is said to have an ROABP of width w* in the order *$\sigma \in S_n$, if there exists a width w ROABP $R(\mathbf{x})$ that computes $f(\mathbf{x})$ in the order σ. We denote the class of such polynomials by* $\mathrm{ROABP}[\sigma](n, d, w)$.
Further, we use $\mathrm{ROABP}[\exists](n, d, w)$ *to denote the class of polynomials that have a width w ROABP in* some *order. That is,* $\mathrm{ROABP}[\exists](n, d, w) = \bigcup_{\sigma \in S_n} \mathrm{ROABP}[\sigma](n, d, w)$.

We can then extend this definition naturally as follows.

▶ **Definition 11** (ROABPs in every order). *For $n, d, w \in \mathbb{N}$, an n-variate polynomial $f(\mathbf{x})$ of* individual *degree d is said to have an ROABP of width w* in every order*, if for all permutations $\sigma \in S_n$, there exists a width w ROABP $R_{(\sigma)}(\mathbf{x})$ that computes $f(\mathbf{x})$ in the order σ.*
We denote this class of polynomials by $\mathrm{ROABP}[\forall](n, d, w)$.

Now, based on the properties of the *coefficient matrices*, we define the two subclasses of ROABPs that Theorem 2 talks about.

▶ **Definition 12** (Commutative ROABPs). *An n-variate, individual degree d ROABP of width w is called a* commutative ROABP *if its coefficient matrices are all $w \times w$ matrices that are (pairwise) commutative.*
We refer of the class of polynomials computed by such ROABPs by $\mathsf{commROABP}(n, d, w)$.

▶ **Definition 13** (Diagonal ROABPs). *An n-variate, individual degree d ROABP of width w is called a* diagonal ROABP *if its coefficient matrices are $w \times w$* diagonal *matrices. We refer of the class of polynomials computed by such ROABPs by* $\mathsf{diagROABP}(n, d, w)$.

Further, we define other concepts about polynomials like *depth-3 powering circuits*, *Waring rank* and *Tensor rank*, since we talk about the connections between them and subclasses of ROABPs defined above.

▶ **Definition 14** (Depth 3 powering circuits ($\Sigma \wedge \Sigma$)). *Over the field $\mathbb{C}$, a depth 3 powering circuit of size s, computes an n-variate, (total) degree d polynomial as an $\mathbb{C}$-linear combination of s terms, each of which is a $\leq d$th power of an $\mathbb{C}$-linear form in the underlying variables $x_1, \ldots, x_n$.*
That is, vectors $\mathbf{a}_1, \ldots, \mathbf{a}_s \in \mathbb{C}^{n+1}$, constants $\beta_1, \ldots, \beta_s$, and $d_1, d_2, \ldots, d_s \in \{0, \ldots, d\}$, define the following n-variate, degree-d, size s depth 3 powering circuit.

$$C(\mathbf{x}) = \sum_{i \in [s]} \beta_i \left(a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n\right)^{d_i}$$

▶ **Definition 15** (Waring rank). *For an n-variate, degree-d polynomial $f(\mathbf{x}) \in \mathbb{C}[\mathbf{x}]$, the* Waring rank *of f is defined to be the size of the smallest depth 3 powering circuit that computes it. We will denote the Waring rank of a polynomial f by* $\mathrm{WR}(f)$.

▶ **Definition 16** (Dimension of partial derivatives). *For an n-variate polynomial $f(\mathbf{x}) \in \mathbb{C}[\mathbf{x}]$, the* dimension of partial derivatives*, which we shall denote by* $\mathrm{DPD}(f)$ *is defined as* $\mathrm{DPD}(f) = \dim\left(\mathrm{span}_{\mathbb{C}}\left\{\partial_{\mathbf{e}} f : \mathbf{e} \in \mathbb{N}^n\right\}\right)$. *Here, $\partial_{\mathbf{e}} f$ denotes the partial derivative $\frac{\partial^{|\mathbf{e}|} f}{\partial x_1^{e_1} \cdots \partial x_n^{e_n}}$.*

Finally, we define the border of diagonal ROABPs as follows, which coincides with the definition of commonly known definition of *border-tensor-rank*.

▶ **Definition 17** (Border of diagonal ROABPs). *For any polynomial $f(\mathbf{x}) \in \mathbb{C}[\mathbf{x}]$, $f(\mathbf{x})$ is in the class $\overline{\mathsf{diagROABP}(n, d, w)}$ if there exists a polynomial $g \in \mathbb{C}(\epsilon)$ in the class* $\mathsf{diagROABP}(n, d, w)$ *such that $f = \lim_{\epsilon \to 0} g$.*

Organization of the paper

The proof the main theorem of this paper(Theorem 2) can be found in Section 4. The appendix is dedicated to studying the algebraic structure of commutative ROABPs, which gives us the necessary ingredients to prove the main theorem. There we first study the "singly-generated" case in Subsection A.1, followed by the structure of general commutative matrix rings in Subsection A.2.

3 Open questions

Owing to the connections of subclasses of ROABPs with other well-studied models, we believe that resolving any of the questions stated in Section 1 in any direction would be very interesting to the algebraic complexity community, and might even lead to new approaches for PIT of ROABPs and depth 3 powering circuits.

A specific follow-up question to our main theorem(Theorem 2) is that of finding an appropriate converse. For example, is it true that if diagonal ROABPs can efficiently simulate commutative ROABPs, then dimension of partial derivatives essentially captures the Waring rank of any polynomial? It is not clear how one would go about proving the above statement directly. For proving the contrapositive, the main technical challenge seems to be to arrive at a candidate commutative ROABP using a polynomial that would witness the separation between dimension of partial derivatives and Waring rank.

4 Proof of the main theorem

We start with an observation about diagonal ROABPs that gives an *equivalent* alternate view of the model, which will be useful for our results.

▶ **Observation 18** (Alternate view of diagonal ROABPs)**.** *If $f(x_1, \ldots, x_n)$ has a* diagonal ROABP *of width w, then there is a polynomial $g(t, \mathbf{x})$ with $\deg_t(g) \leq nw$, such that $f(\mathbf{x}) = \sum_{j\in[w]} g(j, \mathbf{x})$.*

Proof. Suppose $f(\mathbf{x}) = \sum_{j\in[w]} \prod_{i\in[n]} f_{j,i}(x_i)$. Then we define polynomials $L_1(t), \ldots, L_w(t)$ such that for each $j, k \in [w]$, $L_j(k) = 1$ if $j = k$ and $L_j(k) = 0$ otherwise. Such polynomials always exist, and are called *Lagrange basis polynomials*.

For each $i \in [n]$, define $g_i(t, x_i) := \sum_{j\in[w]} L_j \cdot f_{j,i}(x_i)$, and let $g(t, \mathbf{x}) = \prod_{i\in[n]} g_i(t, x_i)$. Then $g(t = j, \mathbf{x}) = \prod_{i\in[n]} f_{j,i}(x_i)$, and hence $f(\mathbf{x}) = \sum_{j\in[w]} g(j, \mathbf{x})$ as required. ◀

4.1 An alternate view of commutative ROABPs

▶ **Definition 19.** *For an ideal $J \subset \mathbb{C}[\mathbf{t}]$, and a $G \in \mathbb{C}[\mathbf{t}, \mathbf{x}]$ given by $G = \sum_{\mathbf{e}} \mathrm{coeff}_{\mathbf{x}^\mathbf{e}}(G)(\mathbf{t}) \cdot \mathbf{x}^\mathbf{e}$, we define the polynomial $\tilde{G} = (G \bmod J)$ as follows.*

$$\tilde{G} := \sum_{\mathbf{e}} (\mathrm{coeff}_{\mathbf{x}^\mathbf{e}}(G)(\mathbf{t}) \bmod J) \cdot \mathbf{x}^\mathbf{e}$$

Here $(g(\mathbf{t}) \bmod J)$ for any $g(\mathbf{t})$ is defined as per Definition 34.

Using the above definition, given any commutative ROABP, we can come up with a product of univariates over $\mathbf{x}$s that is related to it in the following sense.

▶ **Lemma 20.** *Suppose $f(\mathbf{x}) = \mathbf{b}^\intercal \left(\prod_{i\in[n]} \left(A_{i,0} + A_{i,1}x_i + \cdots + A_{i,d}x_i^d \right) \right) \mathbf{c}$, is a commutative-ROABP of width w computing $f(\mathbf{x})$.*

Then there exists an ideal $J \subset \mathbb{C}[t_1, \ldots, t_r]$ with a finite variety, and $G(\mathbf{t}, \mathbf{x}) := \prod_i G_i(\mathbf{t}, x_i)$, such that for $\tilde{G}(\mathbf{t}, \mathbf{x}) := G(\mathbf{t}, \mathbf{x}) \bmod J$, $f(\mathbf{x})$ can be expressed as a linear combination of the $\mathbf{t}$-coefficients of $\tilde{G}$.

Furthermore, $|\mathbf{t}| = r \leq \min\{w^2, n(d+1)\}$ and the $\mathbf{t}$-degree of each G_i is at most w^2.

Proof. Let $F(\mathbf{x})$ denote the $w \times w$ matrix with entries in $\mathbb{C}[\mathbf{x}]$, so that $f(\mathbf{x}) = \mathbf{b}^\intercal F(\mathbf{x})\mathbf{c}$. Let $\mathbf{A} = \{A_1, \ldots, A_r\}$ be such that the ring $\mathbb{C}[A_1, \ldots, A_r]$ is the same as that generated by the coefficient matrices $\{A_{i,j}\}$. It is easy to see that $r \leq \min\{w^2, n(d+1)\}$.

We define the ideal J as follows: $J = \{g(\mathbf{t}) \in \mathbb{C}[\mathbf{t}] : g(A_1, \ldots, A_r) = 0\}$. Let $N_J = \{\mathbf{t}^{\mathbf{a}_1}, \ldots, \mathbf{t}^{\mathbf{a}_m}\}$ be the *normal set* of J; then $|N_J| = m \leq w^2$, as the quotient ring of J is isomorphic to $\mathbb{C}[\mathbf{A}] \subset \mathbb{C}^{w \times w}$ (see Lemma 23). For each i, j let $G_{i,j}(\mathbf{t})$ be the polynomial with monomials from N_J such that $G_{i,j}(\mathbf{A}) = A_{i,j}$. We define $G_i(\mathbf{t}, x_i) = \sum_j G_{i,0}(\mathbf{t}) x_i^j$ for each $i \in [n]$. Since N_J is closed under divisions, the degree of any $\mathbf{t}^{\mathbf{a}} \in N_J$ is at most w^2, and hence $\deg_{\mathbf{t}}(G_i) = \deg(G_{i,j}) \leq w^2$ for all i.

Let $\tilde{G} := (G \bmod J) = \sum_{\mathbf{a} \in N_J} \tilde{g}_{\mathbf{a}}(\mathbf{x}) \mathbf{t}^{\mathbf{a}}$ for some $\tilde{g}_{\mathbf{a}}(\mathbf{x})$s, which we call the "$\mathbf{t}$-coefficients of G".

$$
\begin{aligned}
f(\mathbf{x}) &= \sum_{k,\ell \in [w]} b_k c_\ell \cdot F(\mathbf{x})[k, \ell] \\
\text{(By definition of } G) &= \sum_{k,\ell \in [w]} b_k c_\ell \cdot (G(\mathbf{A}, \mathbf{x}))\,[k, \ell] \\
\text{(By definition of } J) &= \sum_{k,\ell \in [w]} b_k c_\ell \cdot \left(\tilde{G}(\mathbf{A}, \mathbf{x})\right)[k, \ell] \\
\text{(Expanding } \tilde{G}) &= \sum_{k,\ell \in [w]} b_k c_\ell \cdot \left(\sum_{\mathbf{a} \in N_J} \tilde{g}_{\mathbf{a}}(\mathbf{x}) \mathbf{A}^{\mathbf{a}}\right)[k, \ell] \\
\text{(For } A_{\mathbf{a}} = \mathbf{A}^{\mathbf{a}}) &= \sum_{\mathbf{a} \in N_J} \left(\sum_{k,\ell \in [w]} b_k c_\ell A_{\mathbf{a}}[k, \ell]\right) \tilde{g}_{\mathbf{a}}(\mathbf{x}) = \sum_{\mathbf{a} \in N_J} \beta_{\mathbf{a}} \tilde{g}_{\mathbf{a}}(\mathbf{x})
\end{aligned}
$$

In the last line above, $A_{\mathbf{a}} \in \mathbb{F}^{w \times w}$ is the matrix that the "monomial" $\mathbf{A}^{\mathbf{a}}$ evaluates to. ◀

4.2 Evaluating derivatives of polynomials

We now show that for any polynomials $g(\mathbf{t})$, $h(\mathbf{t})$, and any point $\bar{\alpha} \in \mathbb{C}^r$, the value $(D_h(g))(\bar{\alpha})$ can be obtained as a linear combination of $O(d', \mathrm{WR}(h))$ evaluations of the polynomial g, where $d' = \max\{\deg(g), \deg(h)\}$. This is a known fact (see e.g. [16]). We only state the lemma here, and provide a proof in the full version.
We start with a fact about the "symmetry" between $D_h(g)(\bar{0})$ and $D_g(h)(\bar{0})$ that we will need.

▶ **Fact 4.1.** *For any $g, h \in \mathbb{C}[t_1, \ldots, t_r]$, $D_g(h)(\bar{0}) = D_h(g)(\bar{0}) = \sum_{\mathbf{e} \in \mathbb{N}^r} \mathbf{e}! g_{\mathbf{e}} h_{\mathbf{e}}$.*

▶ **Lemma 21** (Functionals and Waring rank)**.** *Let $g, h \in \mathbb{C}[t_1, \ldots, t_r]$ be polynomials of degree at most d', and suppose $\mathrm{WR}(h) \leq s$. Then there exist $W = O(s \cdot d')$ points $\mathbf{y}_1, \ldots, \mathbf{y}_W$ such that $D_h(g)(\bar{0}) = D_g(h)(\bar{0})$ can be expressed as a linear combination of $g(\mathbf{y}_1), \ldots, g(\mathbf{y}_W)$.*

4.3 The proof

We now have all the pieces required to prove the main theorem, which we first restate.

▶ **Theorem 2.** *For any $n, r \in \mathbb{N}$, let $S(r, m)$ denote the smallest $\Sigma \wedge \Sigma$-size required to compute any r-variate polynomial f with $\mathrm{DPD}(f) \leq m$.*
Then for all $n, d, w \in \mathbb{N}$, $\mathsf{commROABP}(n, d, w) \subseteq \mathsf{diagROABP}\left(n, d, S(w^2, w^2) n w^4\right)$.

Proof. Let $F(\mathbf{x}) = \prod_{i=1}^n \left(\sum_{j=0}^d A_{i,j} x_i^j\right)$, and let $f(\mathbf{x}) = \mathbf{b}^\intercal F(\mathbf{x}) \mathbf{c}$ be the corresponding commutative ROABP of width w.

Moving to the polynomial world: From Lemma 20, there is a $G(\mathbf{t}, \mathbf{x}) = \prod_{i\in[n]} G_i(\mathbf{t}, x_i)$ such that $f(\mathbf{x})$ is a linear combination of the $\mathbf{t}$-coefficients of $\tilde{G} := G \bmod J$, where J is the *ideal of dependencies* of the coefficient matrices $\{A_{i,j}\}$.
Let $r = |\mathbf{t}|$, $\mathbf{V}(J) = \{\bar{\alpha}_1, \ldots, \bar{\alpha}_z\}$, and $N_J = \mathrm{NS}(J)$ with $m = |N_J|$. Then $r, m \le w^2$ and $\deg_{\mathbf{t}}(G_i) \le w^2$ for all $i \in [n]$, and there exist $\beta_{\mathbf{a}}$s and $\tilde{g}_{\mathbf{a}}(\mathbf{x})$s such that

$$f(\mathbf{x}) = \sum_{\mathbf{a}\in N_J} \beta_{\mathbf{a}} \tilde{g}_{\mathbf{a}}(\mathbf{x}).$$

Coefficients from derivatives: Next, the results from [12, 13] (Lemma 36) imply that there exist m polynomials $\{h_{u,v}(\mathbf{t})\}$ such that:

- $\mathrm{DPD}(h_{u,v}) \le m$ for all $h_{u,v}$, and
- For any $\mathbf{a} \in N_J$, $\mathrm{coeff}_{\mathbf{a}}(\tilde{G}) = \sum_{u,v} \gamma^{\mathbf{a}}_{u,v} (D_{h_{u,v}}(G))(\bar{\alpha}_u)$, for some $\{\gamma^{\mathbf{a}}_{u,v}\} \subset \mathbb{C}$.

Derivatives using evaluations: Then, using Lemma 21 we see that for any polynomial h with $s := \mathrm{WR}(h)$ and for any polynomial G with $\deg(g), \deg(h) \le d'$, there exist at most $s \cdot d'$ points $\mathbf{y}_1, \ldots, \mathbf{y}_{sd'} \in \mathbb{C}^r$ and constants $\lambda_1, \ldots, \lambda_{sd'} \in \mathbb{C}$ such that :

$$(D_h(G))(\bar{\alpha}) = \sum_{q=1}^{sd'} \lambda_q G(\mathbf{y}_q).$$

Thus, for all u, v, $O(\mathrm{WR}(h_{u,v}) \cdot \max\{\deg_{\mathbf{t}}(G), \deg(h_{u,v})\}) = O(S(r,m) \cdot nw^2)$ evaluations of G are enough to obtain $(D_{h_{u,v}}(G))(\bar{\alpha}_u)$.

Putting everything together: Combining all the steps, we get the following.

$$\begin{aligned}
f(\mathbf{x}) &= \sum_{\mathbf{a}\in N_J} \beta_{\mathbf{a}} \tilde{g}_{\mathbf{a}}(\mathbf{x}) \\
&= \sum_{\mathbf{a}\in N_J} \beta_{\mathbf{a}} \sum_{u,v} \gamma^{\mathbf{a}}_{u,v} (D_{h_{u,v}}(G))(\bar{\alpha}_u) \\
(\text{Rearranging}) &= \sum_{u,v} \left(\sum_{\mathbf{a}\in N_J} \beta_{\mathbf{a}} \gamma^{\mathbf{a}}_{u,v} \right) (D_{h_{u,v}}(G))(\bar{\alpha}_u) \\
(\text{For appropriate } \beta'\text{s}) &= \sum_{u,v} \beta'_{u,v} (D_{h_{u,v}}(G))(\bar{\alpha}_u) \\
\left(\mathrm{DPD}(h_{u,v}) \le m,\ \deg(G) \le nw^2\right) &= \sum_{u,v} \beta'_{u,v} \sum_{q=1}^{S(r,m)\cdot nw^2} \lambda_q G(\mathbf{y}_q, \mathbf{x}) \\
\therefore f(\mathbf{x}) &= \sum_{q'=1}^{m\cdot S(r,m)\cdot nw^2} \mu_{q'} \prod_{i\in[n]} G_i(\mathbf{y}_q, x_i)
\end{aligned}$$

Thus, as $m, r \le w^2$, we get a diagonal ROABP for $f(\mathbf{x})$ of width $O(w^2 \cdot S(w^2, w^2) \cdot nw^2) = O(S(w^2, w^2) \cdot nw^4)$. ◀

▶ Remark 22. Suppose there exists an explicit polynomial f that is computable by a commutative ROABP of polynomial width but any diagonal ROABP computing f requires width super-polynomial in n. Let w be the width of the commutative ROABP, and let J be the ideal of dependencies of its coefficient matrices. By Lemma 36 there exist polynomials $\{h_{u,v}(\mathbf{t})\}$ with $|\mathbf{t}| \le w^2$, such that $\mathrm{DPD}(h_{u,v}) \le w^2$. But if $\mathrm{WR}(h_{u,v}) = \mathrm{poly}(w)$ for each u, v, then we should get a diagonal ROABP of width $\mathrm{poly}(w)$, which is a contradiction. Thus, a separation between commutative and diagonal ROABPs also leads to an explicit polynomial that witnesses the separation dimension of partial derivatives and Waring rank.

References

1. Walter Baur and Volker Strassen. The complexity of partial derivatives. *Theoretical Computer Science*, 22:317–330, 1983. doi:10.1016/0304-3975(83)90110-X.
2. Prerona Chatterjee, Mrinal Kumar, Adrian She, and Ben Lee Volk. A quadratic lower bound for algebraic branching programs. In Shubhangi Saraf, editor, *35th Computational Complexity Conference, CCC 2020, July 28-31, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 169 of *LIPIcs*, pages 2:1–2:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.CCC.2020.2.
3. David A. Cox, John B. Little, and Donal O'Shea. *Ideals, Varieties and Algorithms*. Undergraduate texts in mathematics. Springer, 2007. doi:10.1007/978-0-387-35651-8.
4. David S. Dummit and Richard M. Foote. *Abstract Algebra*. John Wiley and Sons, Inc., second edition, 1999.
5. Pranjal Dutta, Prateek Dwivedi, and Nitin Saxena. Demystifying the border of depth-3 algebraic circuits. In *62nd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2021)*, 2021. URL: https://www.cse.iitk.ac.in/users/nitin/papers/border-depth3.pdf.
6. Michael A. Forbes and Amir Shpilka. Quasipolynomial-time identity testing of non-commutative and read-once oblivious algebraic branching programs. In *54th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2013)*, pages 243–252, 2013. doi:10.1109/FOCS.2013.34.
7. Ignacio García-Marco, Pascal Koiran, Timothée Pecatte, and Stéphan Thomassé. On the complexity of partial derivatives. In Heribert Vollmer and Brigitte Vallée, editors, *34th Symposium on Theoretical Aspects of Computer Science, STACS 2017, March 8-11, 2017, Hannover, Germany*, volume 66 of *LIPIcs*, pages 37:1–37:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017. doi:10.4230/LIPIcs.STACS.2017.37.
8. Rohit Gurjar, Arpita Korwar, and Nitin Saxena. Identity testing for constant-width, and commutative, read-once oblivious abps. *Theory of Computing*, 13(1):1–21, 2017. doi:10.4086/toc.2017.v013a002.
9. Neeraj Kayal. Affine projections of polynomials. In *44th Annual ACM Symposium on Theory of Computing (STOC 2012)*, pages 643–662, 2012. doi:10.1145/2213977.2214036.
10. Mrinal Kumar and Ben Lee Volk. A lower bound on determinantal complexity. In *36th Annual Computational Complexity Conference (CCC 2021)*, volume 200 of *LIPIcs*, pages 4:1–4:12. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.CCC.2021.4.
11. Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. Superpolynomial lower bounds against low-depth algebraic circuits. *Electron. Colloquium Comput. Complex.*, page 81, 2021. URL: https://eccc.weizmann.ac.il/report/2021/081.
12. M.G. Marinari, H.M. Möller, and T. Mora. Gröbner bases of ideals defined by functionals with an application to ideals of projective points. *Applicable Algebra in Engineering, Communication and Computing*, 4(2):103–145, 1993. doi:10.1007/BF01386834.
13. H. Michael Möller and Hans J. Stetter. Multivariate polynomial equations with multiple zeros solved by matrix eigenproblems. *Numerische Mathematik*, 70, 1995. doi:10.1007/s002110050122.
14. Noam Nisan. Lower bounds for non-commutative computation. In *23rd Annual ACM Symposium on Theory of Computing (STOC 1991)*, pages 410–418, 1991. doi:10.1145/103418.103462.
15. Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997. doi:10.1007/BF01294256.
16. Kevin Pratt. Waring rank, parameterized and exact algorithms. In David Zuckerman, editor, *60th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2019)*, pages 806–823. IEEE Computer Society, 2019. doi:10.1109/FOCS.2019.00053.
17. Ran Raz. Elusive functions and lower bounds for arithmetic circuits. *Theory of Computing*, 6(1):135–177, 2010. doi:10.4086/toc.2010.v006a007.

18 Chandan Saha. Factoring Polynomials over Finite Fields using Balance Test. In *25th Symposium on Theoretical Aspects of Computer Science (STACS 2008)*, pages 609–620, 2008.

19 Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity. Github survey, 2015. URL: `https://github.com/dasarpmar/lowerbounds-survey/releases/`.

20 Yaroslav Shitov. How hard is the tensor rank?, 2016. `arXiv:1611.01559`.

21 Amir Shpilka and Avi Wigderson. Depth-3 arithmetic circuits over fields of characteristic zero. *Computational Complexity*, 10(1):1–27, 2001. `doi:10.1007/PL00001609`.

22 V. Strassen. Gaussian Elimination is not Optimal. *Numerische Mathematik*, 13(3):354–356, 1969. `doi:10.1007/BF02165411`.

23 Leslie G. Valiant. Completeness Classes in Algebra. In *11th Annual ACM Symposium on Theory of Computing (STOC 1979)*, pages 249–261, 1979. `doi:10.1145/800135.804419`.

A Algebraic structure of commutative ROABPs

This section is aimed at equipping the reader with the algebraic-geometric concepts about *rings generated by commuting matrices*, that are required to understand the results in [12] and [13] (Lemma 32 and Lemma 36). It is therefore largely expository, and readers who are comfortable with these concepts may skip it.

We start by analysing rings generated by a single matrix in Subsection A.1, and then extend our observations to general rings of matrices in Subsection A.2.

A.1 Rings generated by a single matrix

For any matrix $A \in \mathbb{C}^{w\times w}$, the commutative ring generated by A that is denoted by $\mathbb{C}[A]$, is the set of all matrices that can be written as univariate polynomials in terms of A. In other words, $\mathbb{C}[A] := \{p(A) : p(t) \in \mathbb{C}[t]\}$.

Observe that the matrices $I(= A^0), A, A^2, \ldots, A^w$ satisfy the linear dependency that is given by the *characteristic polynomial of* A: $\det(A - tI) \in \mathbb{C}[t]$. Thus, $\mathbb{C}[A]$ is a vector space (over $\mathbb{C}$) of dimension at most w.

In fact the dimension of $\mathbb{C}[A]$ could be even smaller, and it is captured by the degree of the *minimal polynomial of* A: the smallest degree polynomial $p(t)$ such that $p(A)$ is the zero matrix; and the ideal generated by p, $\langle p\rangle := \{q(t) \in \mathbb{C}[t] : q(t) \text{ is divisible by } p(t)\}$, characterises the ring $\mathbb{C}[A]$. The following fact formalises this relationship.

▶ **Fact A.1.** *Let $A \in \mathbb{C}^{w\times w}$ and let $p(t) \in \mathbb{C}[t]$ be its minimal polynomial. Then the ring generated by A, $\mathbb{C}[A]$, is* isomorphic *to the quotient ring $\mathbb{C}[t]/\langle p\rangle$.*

Proof. Define $\Phi : \mathbb{C}[t] \to \mathbb{C}[A]$ such that $\Phi(q(t)) = q(A)$ for any q. Then the following facts together show that the restriction of Φ on $\mathbb{C}[t]/\langle p\rangle$ is a ring isomorphism by the *first ring isomorphism theorem* (see e.g. [4]).

- Φ is a ring homomorphism: $\Phi(q_1 + q_2 \cdot q_3) = (q_1 + q_2 \cdot q_3)(A) = q_1(A) + q_2(A) \cdot q_3(A)$.
- Φ is *onto*: Trivially follows from the definition of $\mathbb{C}[A]$.
- $\ker \Phi = \langle p\rangle$: Suppose $\Phi(q) = 0$. Then $q(A) = 0$, which implies that $q(t) = p(t) \cdot q'(t)$ as $p(t)$ is the minimal polynomial of A. ◀

Let us now focus on the quotient ring of the ideal generated by an arbitrary polynomial $p(t)$; we shall later rephrase our findings in terms of matrices.

Suppose $p(t) = (t - \alpha_1)^{e_1}(t - \alpha_2)^{e_2} \cdots (t - \alpha_z)^{e_z}$, of degree $m = \sum_u e_u$. Since we are working over $\mathbb{C}$, this is true without loss of generality. Let p_u be the polynomial $(t - \alpha_u)^{e_u}$, for each $u \in [z]$. Then any polynomial $q(t)$ is divisible by p_u whenever α_u is a root of $q(t)$ and its first $(e_u - 1)$ derivatives. In fact, $q(t)$ is divisible by $p = \prod_u p_u$, exactly when the above condition holds for each $u \in [z]$.

▶ **Fact A.2.** *A polynomial $q(t)$ is divisible by $p(t) = \prod_{u\in[z]}(t-\alpha_u)^{e_u}$ if and only if:*

$$\forall u \in [z], \qquad q(\alpha_u) = \frac{\partial q}{\partial t}(\alpha_u) = \frac{\partial^2 q}{\partial t^2}(\alpha_u) = \cdots = \frac{\partial^{e_u-1} q}{\partial t^{e_u-1}}(\alpha_u) = 0.$$

In other words, the $\sum_u e_u = m$ values obtained by evaluating the appropriate derivatives of q at the corresponding roots of p, tell us whether p divides q. These evaluations of derivatives in fact give us some more information about q with respect to the ideal $\langle p \rangle$, which we now see.

Derivatives characterise the quotient ring. For any polynomials $p(t), q(t)$ we define the "remainder polynomial" $q(t)$ mod $p(t)$ as follows.

$$q(t) \bmod p(t) = \tilde{q}(t), \text{ such that } q(t) = q'(t)p(t) + \tilde{q}(t), \text{ with } \deg(\tilde{q}) < \deg(p)$$

Suppose $p(t)$ is a polynomial of degree m, then $\tilde{q}(t)$ is clearly a polynomial of degree at most $m-1$. It turns out that the d evaluations of derivatives of q given in Fact A.2 completely determine $\tilde{q}$.

▶ **Fact A.3.** *Suppose $p(t) = \prod_{u\in[z]}(t-\alpha_u)^{e_u}$ has degree m, then there exist m^2 constants $\{\gamma^a_{u,v}\} \subset \mathbb{C}$ such that for any polynomial $q(t)$, we have*

$$\forall 0 \le a \le m-1, \qquad \tilde{q}_a = \sum_{\substack{u\in[z]\\ v\in[e_u]}} \gamma^{(a)}_{u,v} \cdot \frac{\partial^v q}{\partial t^v}(\alpha_u),$$

where $\tilde{q}(t) := \sum_{0\le j\le m-1} \tilde{q}_j t^j = q(t) \bmod p(t)$.

A.2 General commutative matrix rings

The above observations about "univariate" rings can be summarised as follows. Firstly, any matrix ring is isomorphic to the quotient ring of an ideal, where this ideal contains all polynomial dependencies that the generator matrix satisfies (Fact A.1); thus every matrix in the ring corresponds to a polynomial modulo this ideal.
Secondly, the remainder of any polynomial q with respect to this ideal is completely determined by the evaluations of certain derivatives of q at appropriate points (Fact A.3).

We shall now see the multivariate analogues of the above facts, which tell us about rings generated by multiple commuting matrices.

To fix some notation, suppose that we have been given the $w \times w$ matrices $A_1, \ldots, A_r$ that all commute with each other. These matrices therefore generate a commutative ring of matrices denoted by $\mathbb{C}[A_1, \ldots, A_r]$, whose algebraic properties we shall now provide.

A.2.1 Matrix rings as quotient rings of ideals

Recall that for the ring $\mathbb{C}[A]$, the corresponding ideal was $\langle p(t) \rangle$, where p was the minimal polynomial of A. The ideal $\langle p(t) \rangle$ precisely contains all the polynomials $q(t)$ for which $q(A) = 0$. Therefore a natural choice for the multivariate ideal is the *ideal of dependencies of* $A_1, \ldots, A_r$, $J := \{q(t_1, \ldots, t_r) \in \mathbb{C}[\mathbf{t}] : q(A_1, \ldots, A_r) = 0\}$. Indeed, the quotient ring of J is isomorphic to $\mathbb{C}[A_1, \ldots, A_r]$.

▶ **Lemma 23.** *Suppose $A_1, A_2, \ldots, A_r \in \mathbb{C}^{w\times w}$ are mutually commutative, and let J be their* ideal of dependencies *inside the r-variate polynomial ring $\mathbb{C}[\mathbf{t}]$. Then $\mathbb{C}[A_1, \ldots, A_r]$ is isomorphic to $\mathbb{C}[\mathbf{t}]/J$.*

Proof. Similar to the proof of Fact A.1, we define the map $\Phi : \mathbb{C}[\mathbf{t}] \to \mathbb{C}[A_1, \ldots, A_r]$, which maps $q(\mathbf{t})$ to the matrix $q(A_1, \ldots, A_r)$. This naturally defines the (restricted) map $\phi : {}^{\mathbb{C}[\mathbf{t}]}/_J \to \mathbb{C}[A_1, \ldots, A_r]$, with $\phi(\tilde{q}) = \tilde{q}(\mathbf{A})$.

The following facts are now easy to verify for Φ, which together prove that ϕ is an isomorphism by the *first ring isomorphism theorem* (see e.g. [4]).

- Φ is a ring homomorphism: $\Phi(q_1+q_2\cdot q_3) = (q_1+q_2\cdot q_3)(A_1, \ldots, A_r) = q_1(\mathbf{A})+q_2(\mathbf{A})\cdot q_3(\mathbf{A}) = \Phi(q_1) + \Phi(q_2) \cdot \Phi(q_3)$.
- Φ is *onto*: Trivially follows from the definition of $\mathbb{C}[\mathbf{A}]$.
- $\ker \Phi = J$: Suppose $\Phi(q) = 0$. Then $q(\mathbf{A}) = 0$, which implies that $q(\mathbf{t}) \in J$. ◀

We note an important property of the ideal J, before moving on to the next part. Notice that the minimal polynomials of each of the matrices $A_1, \ldots, A_r$, say $p_1(t_1), p_2(t_2), \ldots, p_r(t_r)$ are elements of J. This means that J contains univariate polynomials in each of its underlying variables. Thus, the set of common zeroes of polynomials in J, also known as the *variety of* J(denoted by $\mathbf{V}(J)$), is finite. One way to see this is that $\mathbf{V}(J) \subseteq \text{roots}(p_1) \times \text{roots}(p_2) \times \cdots \times \text{roots}(p_r)$, where $\text{roots}(p_i)$ denotes the constants in $\mathbb{C}$ where p_i vanishes, and $\times$ denotes the *Cartesian product* of sets. Such ideals are called *zero dimensional ideals*, because their variety is a zero dimensional set in the ambient space $\mathbb{C}^r$.

▶ **Definition 24** (Zero-dimensional ideals). *An ideal $J \subseteq \mathbb{C}[\mathbf{t}]$ is called* zero-dimensional *if its variety is finite; i.e. $|\mathbf{V}(J)| < \infty$.*

A.2.2 Quotient rings of zero dimensional ideals

Since we are interested in zero dimensional ideals J, we shall now assume that $\mathbf{V}(J) = \{v_1, \ldots, v_z\}$ for some $z \in \mathbb{N}$.

Arguably, the statements we have discussed till this point are fairly well-known. But we believe that most of the ideas we shall now see are not as commonly known, especially in the theoretical computer science community. We remark that much of the non-trivial ideas and proofs in this section (Appendix A) belong to previous works [12, 13].

Taking a cue from Fact A.3, for a zero-dimensional ideal J we expect the "multiplicities" of the points in its variety $\mathbf{V}(J)$ to help us find the correct derivatives. In this case, the commonly used definition of multiplicity for multivariate polynomials: multiplicity of w means *all* partial derivatives of order $< w$ vanish, turns out to be a little too coarse. In order to formally introduce the suitable definition, we need the following notion of *derivative operators*, which are like polynomials whose monomials are partial derivatives.

▶ **Definition 25** (Derivative operators). *A* derivative operator *on $\mathbb{C}[t_1, \ldots, t_r]$ is a $\mathbb{C}$-linear combination of* finitely many *partial derivatives of the form $\partial_{\mathbf{a}} : \mathbb{C}[\mathbf{t}] \to \mathbb{C}[\mathbf{t}]$, where $\mathbf{a} \in \mathbb{N}^r$.*

The operator $D = \sum_{\mathbf{a}} \gamma_{\mathbf{a}} \partial_{\mathbf{a}}$ naturally maps a polynomial $q(\mathbf{t}) \in \mathbb{C}[\mathbf{t}]$, to $(\sum_{\mathbf{a}} \gamma_{\mathbf{a}} \cdot \partial_{\mathbf{a}} q(\mathbf{t}))$ which we denote by $D(q)$.

Any polynomial $h(\mathbf{t})$ naturally defines a derivative operator $D_h := \sum_{\mathbf{a} \in \text{supp}(h)} \text{coeff}_h(\mathbf{a}) \partial_{\mathbf{a}}$. Likewise, one can talk about the polynomial that underlies a derivative operator.

In Fact A.3, the set of derivative-evaluations that characterise the ideal generated by a $p = (t - \alpha)^e$, are evaluations at α of derivatives with respect to the monomials $\{t^{e-1}, t^{e-2}, \ldots, t, 1\}$; for multiple factors we take the union of the evaluations for each factor. In particular, there is a "maximum" derivative ${}^{\partial^e}/_{\partial t^e}$, and the other derivatives are obtained by "down-shifting" it (similar to taking all possible derivatives of the underlying monomial). This observation leads us to define the following notion of *shifts* of derivatives and derivative operators.

▶ **Definition 26** (Shifts of derivatives and derivative operators). *For a partial derivative* $\partial_{\mathbf{e}} : \mathbb{C}[\mathbf{t}] \to \mathbb{C}[\mathbf{t}]$ *and a vector* $\mathbf{a} \geq \bar{0}$*, we define the* $\mathbf{a}$-shift of $\partial_{\mathbf{e}}$*, denoted by* $\sigma_{\mathbf{a}}(\partial_{\mathbf{e}})$*, as follows.*

$$\sigma_{\mathbf{a}}(\partial_{\mathbf{e}}) := \begin{cases} \frac{\mathbf{e}!}{(\mathbf{e}-\mathbf{a})!} \cdot \partial_{\mathbf{e}-\mathbf{a}} & \text{if } \mathbf{a} \leq \mathbf{e}, \\ 0 & \text{otherwise.} \end{cases}$$

The definition naturally extends to $\mathbf{a}$-shift of D_h*, denoted by* $\sigma_{\mathbf{a}}(D_h)$*, as follows.*

$$\sigma_{\mathbf{a}}(D_h) := \sum_{\mathbf{e}:\mathbf{e}\geq\mathbf{a}} \text{coeff}_{\mathbf{e}}(h) \cdot \sigma_{\mathbf{a}}\left(\partial_{\mathbf{e}}\right) = \sum_{\mathbf{e}:\mathbf{e}\geq\mathbf{a}} \text{coeff}_{\mathbf{e}}(h) \cdot \frac{\mathbf{e}!}{(\mathbf{e}-\mathbf{a})!} \cdot \partial_{\mathbf{e}-\mathbf{a}}$$

The following observations about derivative operators and their shifts will be useful.

▶ **Observation 27.** *For any derivative operator* D_h *and vector* $\mathbf{a}$*,* $\sigma_{\mathbf{a}}(D_h) = D_{\partial_{\mathbf{a}}(h)}$*.*

▶ **Observation 28.** *For any derivative operator* D_h *and polynomials* $p(\mathbf{t}), q(\mathbf{t})$*, we have the following.*

$$D_h(p \cdot q) = \sum_{\mathbf{a}} \frac{1}{\mathbf{a}!} \cdot \partial_{\mathbf{a}}(p) \cdot \sigma_{\mathbf{a}}(D_h)(q) = \sum_{\mathbf{a}} \frac{1}{\mathbf{a}!} \cdot \partial_{\mathbf{a}}(p) \cdot D_{\partial_{\mathbf{a}}(h)}(q)$$

In the language of shifts of derivative operators, we can say that the set of derivatives with respect to $\left\{t^e, t^{e-1}, \ldots, t, 1\right\}$ is *down-closed*: closed under taking shifts. The following definitions then follow naturally.

▶ **Definition 29** (Down-closed spaces of derivative operators). *A* $\mathbb{C}$*-vector space of derivative operators* Δ *is said to be* down-closed *if for all* $D \in \Delta$*, any shift* D' *of* D*, also belongs to* Δ*.*

▶ **Definition 30** (Closure of an operator). *For a polynomial* $h(t_1, \ldots, t_r) \in \mathbb{C}[\mathbf{t}]$ *and the corresponding derivative operator* D_h*, we define the* closure of D_h *as follows.*

$\Delta(h) := \left\{D_{\partial_{\mathbf{e}}(h)} : \mathbf{e} \in \mathbb{N}^r, \partial_{\mathbf{e}}(h) \neq 0\right\}$.

Ideals with a single point in their variety and closed spaces of derivative operators have the following interesting connection, similar to a univariate ideal $\langle (t-\alpha)^e \rangle$.

▶ **Lemma 31.** *Let* $J \in \mathbb{C}[t_1, \ldots, t_r]$ *be an ideal with* $\mathbf{V}(J) = \{\bar{\alpha}\}$*, then the set* $\Delta(J)$ *of derivative operators defined by* $\Delta(J) := \{D \in \mathbb{C}[\partial t_1, \ldots, \partial t_r] : \forall g \in J, D(g)(\bar{\alpha}) = 0\}$ *a closed vector space.*

Proof. Firstly, for all D_1, D_2, and $\beta \in \mathbb{C}$, $(\beta D_1 + D_2)(f)(\bar{\alpha}) = \beta D_1(f)(\bar{\alpha}) + D_2(f)(\bar{\alpha}) = 0$, just by linearity of differentiation. So $\Delta(J)$ is a vector space over $\mathbb{C}$.

To see that it is closed, suppose $D_h \in \Delta(J)$ for a polynomial $h(\mathbf{t})$, and let $i \in [r]$ be such that the partial derivative $h' := {}^{\partial h}/_{\partial t_i} \neq 0$. Then using Observation 28, for any $g \in J$ we have that $D_h(t_i \cdot g)(\bar{\alpha}) = (t_i \cdot D_h(g) + 1 \cdot D_{h'}(g))(\bar{\alpha}) = v_i \cdot D_h(g)(\bar{\alpha}) + 1 \cdot D_{h'}(g)(\bar{\alpha})$. Now since J is an ideal, $g \in J$ implies that $t_i \cdot g \in I$ and therefore $D_h(t_i \cdot g)(\bar{\alpha}) = 0$; and $D_h(g)(\bar{\alpha}) = 0$ because $g \in J$ and $D_h \in \Delta(J)$. Thus, $D_{h'}(g)(\bar{\alpha}) = 0$ for any $D_h \in \Delta(J)$ and $i \in [r]$ such that ${}^{\partial h}/_{\partial t_i} \neq 0$. The closure under an arbitrary shift $\mathbf{a}$ then follows by induction on the $\mathbf{a}$. ◀

We are now ready to state the following result which follows from the work of Marinari, Möller and Mora [12, Theorem 2.6], which is a suitable multivariate analogue for Fact A.2.

▶ **Lemma 32** (Zero dimensional ideals and derivative operator spaces). *Suppose an ideal* $J \subseteq \mathbb{C}[\mathbf{t}]$ *has variety* $\mathbf{V}(J) = \{\bar{\alpha}_1, \ldots, \bar{\alpha}_z\}$ *and* $\dim_{\mathbb{C}}\left({}^{\mathbb{C}[\mathbf{t}]}/_J\right) = m$*. Then there exist closed spaces of derivative operators* $\Delta_1, \ldots, \Delta_z$ *of dimensions* $m_1, \ldots, m_z$ *with* $\sum_u m_u = m$*, such that for any polynomial* $g(\mathbf{t}) \in \mathbb{C}[\mathbf{t}]$ *we have that* $g \in J$*, if and only if* $\forall u \in [z], \forall D \in \Delta_u : D(g)(\bar{\alpha}_u) = 0$*.*

Thus, every zero-dimensional ideal is characterised by a set of closed spaces of derivative operators, where the number of spaces is equal to the size of the variety. Next, we see how one can obtain "g mod J" given the $\sum_u m_u = m$ derivative-evaluations corresponding to the z bases of $\Delta_1, \ldots, \Delta_z$. To that end, we first formalise what g mod J means and then state a result from [13] that provides the above solution.

A.2.3 Matrices and polynomials in the quotient ring

When dealing with univariate polynomials, it is quite straightforward to define $q(t)$ mod $p(t)$ as $r(t)$, such that $q(t) = q'(t)p(t) + r(t)$ for some polynomial $q'(t)$ with $\deg(r) < \deg(p)$. This is because we intuitively identify $r(t)$ to be "less than" $p(t)$ since it has smaller degree, and thus the concepts of division and remainders extend naturally. However, things are a little more tricky for multivariate polynomials: e.g. which monomial is "smaller"? x^2 or y^2?

We therefore need to fix a consistent way of comparing any two given monomials; we need a *monomial ordering*: a total ordering on monomials that "respects" division/multiplication (see e.g. [3, Chapter 2]). We shall skip the formal definition of a monomial ordering, and just work with the "dictionary ordering" or *lexicographic ordering*: $\mathbf{t}^{\mathbf{a}} \prec \mathbf{t}^{\mathbf{a}'}$ if the smallest $i \in [r]$ with $a_i \neq a'_i$ is such that $a_i < a'_i$. Using the monomial ordering $\prec$, we can define the *leading monomial* of a polynomial, and then *leading monomials of J* for an ideal J.

▶ **Definition 33** (Leading monomials). *For a polynomial $g(\mathbf{t})$, a monomial $\mathbf{t}^{\mathbf{a}} \in \mathrm{supp}(g)$ is said to be the* leading monomial *of g, denoted by $\mathrm{LM}(g)$, if for all $\mathbf{t}^{\mathbf{a}'} \in \mathrm{supp}(g)$ we have that $\mathbf{t}^{\mathbf{a}'} \prec \mathbf{t}^{\mathbf{a}}$.*

Similarly, we define $\mathrm{LM}(J) := \{\mathrm{LM}(g) : g \in J\}$ for an ideal J.

We can then define the remainder of a polynomial with respect to an ideal J.

▶ **Definition 34** (Remainder modulo an ideal). *For a polynomial $g(\mathbf{t})$ and an ideal $J \subset \mathbb{C}[\mathbf{t}]$, we say that $g(\mathbf{t}) \bmod J = \tilde{g}(\mathbf{t})$, if there exist polynomials $g_J(\mathbf{t}) \in J$ and $\tilde{g}(\mathbf{t})$ such that $g(\mathbf{t}) = g_J(\mathbf{t}) + \tilde{g}(\mathbf{t})$, where $\mathrm{LM}(\tilde{g})$ does not belong to the ideal $\langle \mathrm{LM}(J) \rangle$.*

Observe that if $LM(\tilde{g}) \notin \langle \mathrm{LM}(J) \rangle$, then in fact no monomial in $\mathrm{supp}(\tilde{g})$ belongs to the ideal $\langle \mathrm{LM}(J) \rangle$. And thus $\mathrm{supp}(\tilde{g})$ is contained in the "complement of $\langle \mathrm{LM}(J) \rangle$", called the *normal set of J*.

▶ **Definition 35** (Normal set of an ideal). *For an ideal $J \in \mathbb{C}[t_1, \ldots, t_r]$, the* normal set of J *is defined as $\mathrm{NS}(J) := \{\mathbf{t}^{\mathbf{a}} : \mathbf{a} \in \mathbb{N}^r, \mathbf{t}^{\mathbf{a}} \notin \langle \mathrm{LM}(J) \rangle\}$.*

We sometimes overload notation to denote $\mathrm{NS}(J)$ as the set of exponent vectors. That is, $\mathrm{NS}(J) = \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ means $\mathrm{NS}(J) = \{\mathbf{t}^{\mathbf{a}_1}, \ldots, \mathbf{t}^{\mathbf{a}_m}\}$.

Here are some important properties of the normal set of an ideal (see e.g. [12]).

▶ **Fact A.4.** *For any ideal J, its normal set $\mathrm{NS}(J)$ has the following properties.*

- *For any $g(\mathbf{t})$, the polynomial $g \bmod J$ is a linear combination of monomials in $\mathrm{NS}(J)$, and further, $|\mathrm{NS}(J)| = \dim_{\mathbb{C}}(\mathbb{C}[\mathbf{t}]/J)$.*
- *$\mathrm{NS}(J)$ is closed under divisions. That is, if $\mathbf{t}^{\mathbf{a}} \in \mathrm{NS}(J)$ and $\mathbf{t}^{\mathbf{a}'} | \mathbf{t}^{\mathbf{a}}$, then $\mathbf{t}^{\mathbf{a}'} \in \mathrm{NS}(J)$. In particular, $1 \in N_J$ for all ideals J.*

We can now state the result of Möller and Stetter [13] that gives a more explicit version of the correspondence in Lemma 32. The following is a multivariate analogue of Fact A.3.

▶ **Lemma 36** (Consequence of [13, Theorem 1]). *Suppose $J \subset \mathbb{C}[t_1, \ldots, t_r]$ is an ideal with variety $\mathbf{V}(J) = \{\bar{\alpha}_1, \ldots, \bar{\alpha}_z\}$ and normal set $N_J := \mathrm{NS}(J) = \{\mathbf{a}_1, \ldots, \mathbf{a}_w\}$. Let $\Delta_1, \ldots, \Delta_z$ be the characterising derivative operator spaces, with each Δ_u spanned by $\{D_{u,1}, \ldots, D_{u,m_u}\}$, such that $|N_J| = m = \sum_u m_u$.*

Then there exists a set of m^2 constants $\left\{\gamma_{u,v}^{(\mathbf{a})}\right\} \subset \mathbb{C}$, such that for any polynomial $g(\mathbf{t}) \in \mathbb{C}[\mathbf{t}]$ and $\tilde{g}(\mathbf{t}) := (g(\mathbf{t}) \bmod J)$, we have $\mathrm{coeff}_{\mathbf{a}}(\tilde{g}) = \sum_{u,v} \gamma_{u,v}^{(\mathbf{a})}(D_{u,v}(g))(\bar{\alpha}_u)$ for all $\mathbf{a} \in N_J$.